

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:47:14 ; Search time 40 Seconds  
(without alignments)  
1851.527 Million cell updates/sec

Title: US-09-869-677A-2  
Perfect score: 1475  
Sequence: 1 STGAKTAKSDKLKVATNS.....PGDSYVAMKMKNDKISEGL 287

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 54484

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	3.9	40	9 Q9MBX7	Q9mbx7 streptococ
2	56	3.8	44	16 Q25198	Q25198 helicobacte
3	56	3.8	57	10 Q8GZ03	Q8GZ03 arabidopsi
4	55.5	3.8	54	2 Q8GCM8	Q8GCM8 mycoplasma
5	55	3.7	56	17 Q27979	Q27979 archaeoglob
6	53	3.6	58	16 Q97H22	Q97H22 clostridium
7	52	3.5	51	10 Q9A550	Q9A550 oryza sativ
8	51	3.5	56	2 Q8RN73	Q8RN73 streptococ
9	50	3.4	51	16 Q92GS8	Q92GS8 rickettsia
10	50	3.4	54	2 Q9ZND7	Q9ZND7 thermus the
11	49.5	3.4	46	12 Q89058	Q89058 vesicular s
12	49.5	3.4	48	2 Q51197	Q51197 neisseria m
13	49.5	3.4	49	13 Q8UH25	Q8UH25 lavinia exl
14	49.5	3.4	49	13 Q8UH26	Q8UH26 hesperoleuc
15	49.5	3.4	56	15 Q56197	Q56197 human immun
16	49	3.3	49	6 Q9T515	Q9T515 canis famli

17	49	3.3	50	16 Q8K6V1	Q8K6V1 streptococ
18	49	3.3	52	5 Q8ISR4	Q8ISR4 spodiopetra
19	49	3.3	54	9 Q38280	Q38280 lactococcus
20	49	3.3	54	16 Q8BSV2	Q8BSV2 thermoaer
21	48.5	3.3	46	12 Q86121	Q86121 vesicular s
22	48.5	3.3	57	9 Q8WSW3	Q8WSW3 bacterioph
23	48	3.3	53	10 Q9ZSF7	Q9ZSF7 lycium ande
24	48	3.3	55	16 Q9HZW4	Q9HZW4 pseudomonas
25	47.5	3.2	47	2 Q51199	Q51199 neisseria m
26	47.5	3.2	57	16 Q928B3	Q928B3 listeria in
27	47.5	3.2	58	5 Q25629	Q25629 onchocerca
28	47.5	3.2	59	16 Q50881	Q50881 borrelia bu
29	47.5	3.2	60	7 Q9XR55	Q9XR55 oncorhynch
30	47	3.2	45	6 Q8MWV7	Q8MWV7 sus scrofa
31	47	3.2	49	2 Q97J39	Q97J39 neisseria m
32	47	3.2	56	2 Q52223	Q52223 leuconostoc
33	47	3.2	60	16 Q9PMW4	Q9PMW4 campylobact
34	46.5	3.2	50	5 Q8N064	Q8N064 plasmodium
35	46.5	3.2	56	12 Q99H59	Q99H59 human picob
36	46	3.1	50	4 Q14847	Q14847 homo sapien
37	46	3.1	54	9 Q38258	Q38258 lactococcus
38	46	3.1	59	16 Q9KQX1	Q9KQX1 vibrio chol
39	45.5	3.1	46	12 Q89033	Q89033 vesicular s
40	45.5	3.1	55	12 P87567	P87567 canine aden
41	45.5	3.1	56	16 Q8XLR8	Q8XLR8 clostridium
42	45.5	3.1	56	16 Q8RCH4	Q8RCH4 thermoaer
43	45.5	3.1	58	12 Q89032	Q89032 vesicular s
44	45.5	3.1	58	16 Q8RCY3	Q8RCY3 thermoaer
45	45.5	3.1	60	2 Q8VUR2	Q8VUR2 vibrio para

## ALIGNMENTS

RESULT 1					
Q9MBX7	PRELIMINARY;	PRT;	40 AA.		
ID Q9MBX7					
AC Q9MBX7;					
DT 01-OCT-2000 (TREMBLrel. 15, Created)					
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)					
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE Gp40.					
OS Streptococcus thermophilus bacteriophage Sfil18, and					
OC Streptococcus thermophilus bacteriophage Sfil1.					
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.					
OX NCBI_TaxID=74382, 78541;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC SPECIES=Various species;					
RX MEDLINE=99412383; PubMed=10482618;					
RA Lucchini S., Desiere F., Brusow H.;					
RT "Comparative genomics of Streptococcus thermophilus phage species					
RT supports a modular evolution theory.";					
RL J. Virol. 73:8647-8656(1999).					
RN [2]					
RP SEQUENCE FROM N.A.					
RC SPECIES=Streptococcus thermophilus bacteriophage Sfil1;					
RA Desiere F., Lucchini S., Brusow H.;					
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AF158601; AAF63081.1; -					
DR EMBL; AF158600; AAF63053.1; -					
SQ SEQUENCE 40 AA; 5039 MW; 586A9FB508B0B0 CRC64;					
Query Match	3.9%	Score 58;	DB 9;	Length 40;	
Best Local Similarity	34.9%	Pred. No. 2.3e+03;			
Matches 15;	Conservative	9;	Mismatches 11;	Indels 8;	Gaps 2;
Qy	132 KNIAKQIIADPPKKE---TYEKNTKAYAKLEKLDKEKSK 170				
Db	2 KNLFKWIILADEKQKRWVTPYENKCKY---EETIKQLSMK 40				
RESULT 2					

```

025198
ID 025198 PRELIMINARY; PRT; 44 AA.
AC 025198:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein HP0450.
GN HP0450.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kertavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547 (1997).
DR EMBL; AE000560; AAD07522.1; -.
DR TIGR; HP0450; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 44 AA; 5335 MW; 4D2C88A9D9792E82 CRC64;

Query Match 3.8%; Score 56; DB 16; Length 44;
Best Local Similarity 37.5%; Pred. No. 3.6e+03;
Matches 12; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 150 ENLKAAYAKLEKDEKSKSDALAEKGLI 181
DB 13 ETELDRVAELDRKREKQLKIDAPKQEKFI 44

RESULT 3
08G203 PRELIMINARY; PRT; 57 AA.
AC 08G203:
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative heat shock protein.
GN AT3G08910/T1601.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinzaki K.;
RT "Arabidopsis thaliana full-length cDNA."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK117285; BAC41957.1; -.
KW Heat shock.
SQ SEQUENCE 57 AA; 6746 MW; DEECDS9BD4794CE7 CRC64;

Query Match 3.8%; Score 56; DB 10; Length 57;
Best Local Similarity 31.0%; Pred. No. 5e+03;
Matches 22; Conservative 6; Mismatches 27; Indels 16; Gaps 2;

QY 127 GIIYSKIIAKQIIAKDPKPKETEKRLKAYVAKLEKDEKSKSDALAEKGLI 186

```

```

DB 2 GVDYKVLQVDRNKKDDDKKAYRK-----LAKMWHDPKPNKKDA-----EA 45
QY 187 CFKYSKAYGV 197
DB 46 KFKQISEAYDV 56

RESULT 4
08G208 PRELIMINARY; PRT; 54 AA.
AC 08G208:
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ICEF-1A ORF10.
OS Mycoplasma fermentans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2115;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG18;
RX MEDLINE=22336318; PubMed=12446643;
RA Calcutt M.J., Lewis M.S., Wise K.S.;
RT "Molecular Genetic Analysis of ICEF, an Integrative Conjugal Element
RT That Is Present as a Repetitive Sequence in the Chromosome of
RT Mycoplasma fermentans PG18."
RL J. Bacteriol. 184:6929-6941 (2002).
DR EMBL; AY168953; AAN85220.1; -.
SQ SEQUENCE 54 AA; 6541 MW; 89AC9E8A82595854 CRC64;

Query Match 3.8%; Score 55.5; DB 2; Length 54;
Best Local Similarity 28.3%; Pred. No. 5e+03;
Matches 17; Conservative 11; Mismatches 25; Indels 7; Gaps 2;

QY 134 IAKQIIAKDPKPKETEKRLKAYVAKLEKDEKSKSDALAEKGLI 193
DB 2 ISKRVDELKONGIKSLAKYTLERFKMEPEKYEKFRK-ESNKKI-----CNKLFAX 54

RESULT 5
027979 PRELIMINARY; PRT; 56 AA.
AC 027979:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein AF2305.
GN AF2305.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Richardson D.L., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Fleischmann R.D., Kertavage A.R., Graham D.E., Fliedman N.C.,
RA Kirness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulfate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
DR EMBL; AE000945; AAB88955.1; -.
DR TIGR; AF2305; -.

```

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 56 AA; 6986 MW; DBDA220C268A7DF6 CRC64;

Query Match 3.7%; Score 55; DB 17; Length 56;  
Best Local Similarity 32.6%; Pred. No. 5.7e+03;  
Matches 15; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

QY 121 WNLNENGIYSKNIAQIADPNKREYKKNLKYAKLEKDK 166  
DB 7 WLFRTYPIIDYKSGVWNPDLAYKKDDAERYFLAERFVKELEKPLEE 52

## RESULT 6

Q97HZ2 PRELIMINARY; PRT; 58 AA.  
AC Q97HZ2;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DE Hypothetical protein CAC1864.  
GN CAC1864.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_Taxid=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; Pubmed=11466286;  
RA Noelling J., Bleton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hilti J., Wolf Y.I.,  
RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Souaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.,  
RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum."  
RL J. Bacteriol. 183:4823-4838 (2001).  
DR EMBL; AE007695; AAK79828.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 58 AA; 6901 MW; AA4004BC4B721D72 CRC64;

Query Match 3.6%; Score 53; DB 16; Length 58;  
Best Local Similarity 31.6%; Pred. No. 8.1e+03;  
Matches 18; Conservative 10; Mismatches 25; Indels 4; Gaps 2;

QY 131 SKNIAQIADPNKREYKKNLKYAKLEKDKREKSKPDALANKKLIYTS 185  
DB 4 NKNHILKVVYPTAENEAHYKRAELFYISYIKKKIDSEINKF--IIEELKAMDSSK 58

## RESULT 7

Q9AS50 PRELIMINARY; PRT; 51 AA.  
AC Q9AS50;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE P0493G01.20 protein (B1039D07.1 protein).  
GN P0493G01.20 OR B1039D07.1.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaeae; Oryza.  
OX NCBI\_Taxid=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC  
clone: P0493G01.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;

RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, BAC  
clone: B1039D07.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002914; BAB3938.1; -  
DR EMBL; AP003198; BAB64039.1; -  
DR Gramene; Q9AS50; -

SQ SEQUENCE 51 AA; 5649 MW; 5FD4B7CD215874A4 CRC64;

Query Match 3.5%; Score 52; DB 10; Length 51;  
Best Local Similarity 26.5%; Pred. No. 8e+03;  
Matches 9; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

QY 108 EGASEKGEKDPHAMLNLNENGIYSKNIAQIADPNK 141  
DB 4 EGGESEGEVSSMRTTSGILYEEVTRDMTAE 37

## RESULT 8

Q8RN73 PRELIMINARY; PRT; 56 AA.  
AC Q8RN73;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE M. protein (Fragment).  
GN ERM.  
OS Streptococcus equisimilis.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_Taxid=119602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NSR131;  
RA Pimatanolai N., Orataiwan P., Nuntaphad P.;  
RT "Novel emm gene of GCS.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF485834; ALN87710.1; -  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 56 AA; 6416 MW; 544E6E4A64A3122D CRC64;

Query Match 3.5%; Score 51; DB 2; Length 56;  
Best Local Similarity 37.0%; Pred. No. 1e+04;  
Matches 20; Conservative 9; Mismatches 15; Indels 10; Gaps 3;

QY 131 SKNIAQIADPNKREYKKNLKYAKLEKDKREKSKPDALANKKLIYTS 184  
DB 12 SREVTQVLAKMKMK-----LDA--AKLE--NEKDKDIESLKEAIKSYVS 55

## RESULT 9

Q92GS8 PRELIMINARY; PRT; 51 AA.  
AC Q92GS8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical protein RC1044.  
GN RC1044.  
OS Rickettsia conorii.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_Taxid=781;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Malish 7;  
RX MEDLINE=21442074; Pubmed=11557893;  
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RA Raoult D.;  
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098 (2001).

DR EMBL; AE008655; AAL03582.1; -  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 51 AA; 6180 MW; 190F8A6A0F35F4AF CRC64;  
Query Match 3.4%; Score 50; DB 16; Length 51;  
Best Local Similarity 26.9%; Pred. No. 1.1e+04;  
Matches 14; Conservative 9; Mismatches 21; Indels 8; Gaps 2;  
QY 86 NAQKTQKQYFAVSDGIDVYLEGASEKGEDPHALNLNENGIYSKNIYAKQ 137  
DB 3 NRLAEKNKQFNTINIDIVY-----KKQKVDNSIWFNTINDV--NKIPIKQ 46  
RESULT 10  
Q92ND7 PRELIMINARY; PRT; 54 AA.  
AC Q92ND7;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Orf.  
GN ORF.  
OS Thermus thermophilus.  
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
OC Thermus.  
OX NCBI\_TaxID=274;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HB27;  
RA Kobashi N., Nishiyama M., Tanokura M.;  
RT "Aspartate kinase-independent lysine synthesis in an extremely  
RT thermophilic bacterium Thermus thermophilus: lysine is synthesized not  
RT via diaminopimelic acid but a-aminoadipic acid."  
RL EMBL; AB017109; BAA74765.1; -  
DR InterPro; IPR005906; IYGM.  
DR TIGRFAMs; TIGR01206; IYGM; 1.  
SQ SEQUENCE 54 AA; 5812 MW; 835CEB60AF8C52DE CRC64;  
Query Match 3.4%; Score 50; DB 2; Length 54;  
Best Local Similarity 56.2%; Pred. No. 1.2e+04;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 42 IGDPHEPERLPEDAE 57  
DB 35 VGLDPLRLIEPAEEAE 50  
RESULT 11  
Q89058 PRELIMINARY; PRT; 46 AA.  
AC Q89058;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE (VSV) (Strain New Jersey/W62-3) RNA, 3' end (Fragment).  
OS Vesicular stomatitis virus.  
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
OC Rhabdoviridae; Vesiculovirus.  
OX NCBI\_TaxID=11276;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87061240; PubMed=3023698;  
RA Nichol S.T., Holland J.J.;  
RT "genome RNA terminus conservation and diversity among  
RT vesiculoviruses."  
RL J. Virol. 61:200-205(1987).  
DR EMBL; M14938; AAA48490.1; -  
DR InterPro; IPR000448; Rhabd\_nucleocap.  
DR Pfam; PF00945; Rhabd\_ncap; 1.  
DR ProDom; PD002087; Rhabd\_nucleocap; 1.  
FT NON\_TER 46  
SQ SEQUENCE 46 AA; 5221 MW; B9ED65D46145CA5E CRC64;

Query Match 3.4%; Score 49.5; DB 12; Length 46;  
Best Local Similarity 32.6%; Pred. No. 1e+04;  
Matches 15; Conservative 8; Mismatches 18; Indels 5; Gaps 3;  
QY 22 IADMTKAIAGDKIDLHSHIVPIGDPHEPEPLPEDAKTSNADYIFY 67  
DB 1 MAPTVKRIINDSI-IQPKLPANEDPVEY---PADYFK-NNTNIVLY 41  
RESULT 12  
Q51197 PRELIMINARY; PRT; 48 AA.  
AC Q51197;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Class 1 outer membrane protein variable region 2 (Fragment).  
GN PORA.  
OS Neisseria meningitidis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=487;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC105;  
RA McGuinness B.T., Lambden P.R., Heckels J.E.;  
RT "Class 1 outer membrane protein of Neisseria meningitidis: Epitope  
RT analysis of the antigenic diversity between strains, implications for  
RT subtype definition and molecular epidemiology."  
RL Submitted (AUG-1992) to the EMBL/Genbank/DBD databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC105;  
RX MEDLINE=90278344; PubMed=1693651;  
RA McGuinness B.T., Barlow A.K., Clarke I.N., Farley J.E., Anilions A.,  
RA Poolman J.T., Heckels J.E.;  
RT "Purified amino acid sequences of class 1 protein (pOR) from three  
RT strains of Neisseria meningitidis: synthetic peptides define the  
RT epitopes responsible for serosubtype specificity."  
RL J. Exp. Med. 171:1871-1882(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC105;  
RX MEDLINE=89343617; PubMed=2503673;  
RA Barlow A.K., Heckels J.E., Clarke I.N.;  
RT "The class 1 outer membrane protein of Neisseria meningitidis: gene  
RT sequence, structural and immunological similarities to gonococcal  
RT porins."  
RL Mol. Microbiol. 3:131-139(1989).  
DR EMBL; Z14276; CAA78643.1; -  
FT NON\_TER 48  
FT NON\_TER 1  
SQ SEQUENCE 48 AA; 4993 MW; 2CDE167412PDAE71 CRC64;  
Query Match 3.4%; Score 49.5; DB 2; Length 48;  
Best Local Similarity 38.5%; Pred. No. 1.1e+04;  
Matches 15; Conservative 6; Mismatches 11; Indels 7; Gaps 3;  
QY 246 SKDSGIPYSEIFDTSIAKK-----GKPG-DSYVAMMKW 278  
DB 8 SKSAYTPAYVAV-ENGVAKKVAAVGKPGSDVYVAGLNY 45  
RESULT 13  
Q80H25 PRELIMINARY; PRT; 49 AA.  
AC Q80H25;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Triose phosphate isomerase B (Fragment).  
GN TPI-B.

```
OS Lavinia exilicauda (hitch).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Lavinia.
OX NCBI_TaxID=71765;
RN [1]
RP SEQUENCE FROM N.A.
RA Jones W.J., Pettis J., Bernardi G.;
RT "Integrative hybridization in Lavinia."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059467; AAL26707.1.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR PROSITE; PS00171; TIM; 1.
KW Isomerase.
FT NON_TER 1 1
FT NON_TER 49 49
SQ SEQUENCE 49 AA; 5514 MW; A79A533636251894 CRC64;

Query Match 3.4%; Score 49.5; DB 13; Length 49;
Best Local Similarity 31.7%; Pred. No. 1.1e+04;
Matches 19; Conservative 10; Mismatches 14; Indels 17; Gaps 4;

QY 167 AKSKPDIAENKULIVTSEGCFFKYSKAYGVPSAY--IWEINTEBEGTPDQISSLIETLK 224
Db 4 AQTGF--IADN-----VKDWSK---VLAIEPVMALGTGKTASPOQAGVHDKLR 48

RESULT 14
Q8UH26 PRELIMINARY; PRT; 49 AA.
AC O8UH26;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Triose phosphate isomerase B (Fragment).
RT TPI-B.
OS Hesperolencus symmetricus (California roach).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Hesperolencus.
OX NCBI_TaxID=164605;
RN [1]
RP SEQUENCE FROM N.A.
RA Jones W.J., Pettis J., Bernardi G.;
RT "Integrative hybridization in Lavinia."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059466; AAL26706.1.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR PROSITE; PS00171; TIM; 1.
KW Isomerase.
FT NON_TER 1 1
FT NON_TER 49 49
SQ SEQUENCE 49 AA; 5514 MW; A79A533636251894 CRC64;

Query Match 3.4%; Score 49.5; DB 13; Length 49;
Best Local Similarity 31.7%; Pred. No. 1.1e+04;
Matches 19; Conservative 10; Mismatches 14; Indels 17; Gaps 4;

QY 167 AKSKPDIAENKULIVTSEGCFFKYSKAYGVPSAY--IWEINTEBEGTPDQISSLIETLK 224
Db 4 AQTGF--IADN-----VKDWSK---VLAIEPVMALGTGKTASPOQAGVHDKLR 48

RESULT 15
ID 056197 PRELIMINARY; PRT; 56 AA.
AC 056197;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
```

```
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN Env.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98127656; PubMed=9468138;
RA Zhu T., Kober B.T., Nahmias A.J., Hooper E., Sharp P.M., Ho D.D.;
RT "An African HIV-1 sequence from 1959 and implications for the origin
RT of the epidemic."
RL Nature 391:594-597(1998).
DR EMBL; AF030669; AAC09207.1.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1 1
FT NON_TER 56 56
SQ SEQUENCE 56 AA; 6267 MW; 88CB9D41A8263FFB CRC64;

Query Match 3.4%; Score 49.5; DB 15; Length 56;
Best Local Similarity 28.8%; Pred. No. 1.3e+04;
Matches 15; Conservative 8; Mismatches 28; Indels 1; Gaps 1;

QY 77 QAWFTKLVKNAQKTKNDYFVAVSDGIDIVYIEGASKEGKEDPHAMNLNENGI 128
Db 1 QYWSQELKKSANSLNATPAIVAGETDRV-LEVLQRAGRALILHPTRIQGL 51
```

Search completed: February 11, 2004, 16:51:25  
Job time : 63 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:47:44 ; Search time 21 Seconds  
(without alignments)  
1314.305 Million cell updates/sec

Title: US-09-869-677A-2

Perfect score: 1475

Sequence: 1 SSTGAKTAKSDKLKVVATNS.....PGDSYVAMKMNLDKISEGL 287

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 15918

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58.5	4.0	50	2 S02174	anti-epilepsy prot
2	56	3.8	44	2 B64576	hypothetical prote
3	55.5	3.8	33	2 B56635	tubulin alpha chai
4	55	3.7	56	2 A69538	conserved hypothe
5	53	3.6	58	2 A97130	hypothetical prote
6	50	3.4	51	2 D97830	hypothetical prote
7	50	3.4	52	2 C64036	hypothetical prote
8	49.5	3.4	60	2 T39957	degradative enzyme
9	48	3.3	55	2 A83286	hypothetical prote
10	47.5	3.2	57	2 AC1756	hypothetical prote
11	47.5	3.2	59	2 B70243	hypothetical prote
12	47	3.2	48	2 B69606	spore coat protein
13	47	3.2	60	2 D81277	hypothetical prote
14	46.5	3.2	58	2 C46743	conserved hypothe
15	46	3.1	59	2 B82146	cornal keratan su
16	45.5	3.1	59	2 C36493	H+-transporting tw
17	45	3.1	34	2 A28506	pyruvate kinase (E
18	45	3.1	58	2 A35416	DNA topoisomerase
19	45	3.1	59	2 T03335	gene e32 protein -
20	44	3.0	36	2 C69327	hypothetical prote
21	44	3.0	42	2 T13295	hypothetical prote
22	44	3.0	48	2 S26102	outer membrane pro
23	44	3.0	49	2 S57266	ribosomal protein
24	44	3.0	49	2 D82854	hypothetical prote
25	43.5	2.9	44	2 A60155	polygalacturonase
26	43.5	2.9	52	2 B81392	50S ribosomal prot
27	43.5	2.9	58	2 A71527	probable S21 ribos
28	43.5	2.9	58	2 E81683	ribosomal protein
29	43	2.9	26	2 PQ0105	alkaline serine pr

30	43	2.9	38	2 A82450	hypothetical prote
31	43	2.9	46	2 C90562	hypothetical prote
32	43	2.9	49	2 T12136	leghemoglobin - fa
33	43	2.9	50	2 T51287	calcium-binding pr
34	43	2.9	50	2 A54532	probable glycophor
35	43	2.9	54	2 S34728	ribose-phosphate d
36	43	2.9	55	2 A24633	M protein - Strept
37	43	2.9	60	2 A24626	Ig lambda chain V-
38	42.5	2.9	45	2 A83765	hypothetical prote
39	42.5	2.9	51	2 A56371	collagen alpha 1(X
40	42.5	2.9	52	2 E8111	hypothetical prote
41	42.5	2.9	56	2 AG1732	hypothetical prote
42	42.5	2.9	57	2 I68811	MHC class II histo
43	42.5	2.9	58	2 A86495	S21 ribosomal prot
44	42.5	2.9	58	2 F72128	ribosomal protein
45	42.5	2.9	59	2 C70149	ribosomal protein

#### ALIGNMENTS

##### RESULT 1

S02174 anti-epilepsy protein - Manchurian scorpion (fragment)

C:Species: Mesobuthus martenisi (Manchurian scorpion)

C>Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 31-Dec-1993

C:Accession: S02174

R:Zhou, X.H., Yang, D., Zhang, J.H., Liu, C.M., Lei, K.J.

Biochem. J. 257, 509-517, 1989

A:Title: Purification and N-terminal partial sequence of anti-epilepsy peptide from ven

A:Reference number: S02174; MWID:8913444; PMID:2930463

A:Accession: S02174

A:Molecule type: protein

A:Residues: 1-50 <ZHO>

C:Superfamily: scorpion neurotoxin

Query Match 4.0%; Score 58.5; DB 2; Length 50;

Best Local Similarity 34.1%; Pred. No. 8.1e+02;

Matches 15; Conservative 9; Mismatches 15; Indels 5; Gaps 3;

QY 175 AENKKL-IVTSECFYFSKAYGVSPAYIWEINTER--EGTDP 214

DB 7 SDNCKVSCILGNENCKNE-CRAYGASGYCWTVKLAQDCGLPD 49

##### RESULT 2

B64576 hypothetical protein HP0450 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C:Accession: B64576

R:Tomb, J.F., White, O., Kervilange, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D.

Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKen-

son, J.D., Kelley, J.M., Cotton, M.D., Weldman, J.M., Fujii, C., Bowman, C., Wathey, L

Nature 388, 539-547, 1997

A:Authors: Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MWID:97394467; PMID:9252185

A:Accession: B64576

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 144 <TOM>

A:Cross-references: GB:AE000560; GB:AE000511; NID:g2313554; PIDN:ADD07522.1; PID:g23135

Query Match 3.8%; Score 56; DB 2; Length 44;

Best Local Similarity 37.5%; Pred. No. 1e+03;

Matches 12; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 150 EKULKAYVAKLEKLDKYEAKSKFDALANKLI 181

DB 13 ETELDRKVAVELDKREKQLRKDIDAFKQEKKEFI 44

RESULT 3  
B56635  
tubulin alpha chain brain-specific isoctype (clone pTUB9) - chum salmon (fragment)  
C:Species: Oncorhynchus keta (chum salmon)  
C:Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 02-Jul-1998  
C:Accession: B56635  
R:CoE, I.R.; Munro, R.; Sherwood, N.M.  
DNA Seq. 3, 257-262, 1992  
A:Title: Isolation of different brain-specific isoatypes of alpha-tubulins from chum salm  
A:Reference number: A56635; MUID:93208376; PMID:1296820  
A:Contents: brain  
A:Accession: B56635  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-33 <COB>  
A:Note: sequence inconsistent with nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:128390, NCBIPI:128389)  
C:Superfamily: tubulin

Query Match 3.8%; Score 55.5; DB 2; Length 33;  
Best Local Similarity 35.3%; Pred. No. 7.4e+02;  
Matches 12; Conservative 6; Mismatches .11; Indels 5; Gaps 1;  
OY 239 RRMERVKSDSGPIYSIFITDSIAKKRPGDSY 272  
DB 5 REDMALEKRD-----YEEVGTDSIGREAEGRREY 33

RESULT 4  
A69538  
conserved hypothetical protein AF2305 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: A69538  
R:Kleuk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 350, 364-370, 1997  
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: A69538  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-56 <KLE>  
A:Cross-references: GB:AEO00945; GB:AEO00782; NID:92689268; PIDN:AB88955.1; PID:9264821

Query Match 3.7%; Score 55; DB 2; Length 56;  
Best Local Similarity 32.6%; Pred. No. 1.6e+03;  
Matches 15; Conservative 7; Mismatches 24; Indels 0; Gaps 0;  
OY 121 WLNLNGITYSKNTAKQILAKDPKNTYKRLKAVYAKLEKLDKE 166  
DB 7 WLRTTRPIDYSKGVNPLDAYKKDADRYFLAERFVKELEKLE 52

RESULT 5  
A97130  
hypothetical protein CAC1864 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: A97130  
R:Noelling, J.; Bretton, G.; Omoichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: A97130  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-58 <KUR>  
A:Cross-references: GB:AEO01437; PIDN:AAK79828.1; PID:915024842; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1864

Query Match 3.6%; Score 53; DB 2; Length 58;  
Best Local Similarity 31.6%; Pred. No. 2.3e+03;  
Matches 18; Conservative 10; Mismatches 25; Indels 4; Gaps 2;  
OY 131 SKNLAKQILAKDPKNTYKRLKAVYAKL--EKLDKAKSKPDALAKNKLIIVSE 185  
DB 4 NKNLHLKVVYPTALNEAEVHKRSLYFISYIKKLDSEIRNKF--IEELKKAMDSXK 58

RESULT 6  
D97830  
hypothetical protein RC1044 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: D97830  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: D97830  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-51 <KUR>  
A:Cross-references: GB:AEO06914; PIDN:AA03582.1; PID:915620162; GSPDB:GN00173  
A:Gene: RC1044

Query Match 3.4%; Score 50; DB 2; Length 51;  
Best Local Similarity 26.9%; Pred. No. 3.1e+03;  
Matches 14; Conservative 9; Mismatches 21; Indels 8; Gaps 2;  
OY 86 NQKTKNDYFVSDIGIVITLGGSEKQKDPHAKMLNGLITYSKNTAKQ 137  
DB 3 NPLAENKDFENYINNTIDTV-----KKQKWDSDIWFNINDHV--NKIPKQ 46

RESULT 7  
C64036  
hypothetical protein H1562 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: C64036  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,  
.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: C64036  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-52 <TIG>  
A:Cross-references: GB:U32811; GB:L42023; NID:91574407; PIDN:AA023220.1; PID:91574414;  
C:Genetics:  
A:Start codon: GTG

Query Match 3.4%; Score 50; DB 2; Length 52;  
Best Local Similarity 28.3%; Pred. No. 3.2e+03;  
Matches 15; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

OY 138 LIADPKNTYKRLKAVYAKLEKLDKAKSKPDALAKNKLIIVSECPFY 190  
DB 1 MLKSKDK-----VLKLGLEKDKSKAKKVPFGDAD-----LKSQSCDKY 41

RESULT 8  
I39957  
degradative enzyme production factor degr - Bacillus subtilis



C:Species: Bacillus subtilis  
C.Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 15-Oct-1999  
C.Accession: I39957; I39958; A69614  
R:Nagamai, Y.; Tanaka, T.  
J. Bacteriol. 166, 20-28, 1986  
A.Title: Molecular cloning and nucleotide sequence of a DNA fragment from Bacillus natto  
A.Reference number: I39957; MUID:86168015; PMID:3082853  
A.Accession: I39957  
A.Status: preliminary; translated from GB/EMBL/DDBJ  
A.Molecule type: DNA  
A.Residues: 1-60 <RES>  
A.Cross-references: GB:M12917; NID:g143354; PIDN:AAA22671.1; PID:g143355  
R.Yang, M.; Shimotsu, H.; Ferrari, E.; Henner, D.J.  
J. Bacteriol. 169, 434-437, 1987  
A>Title: Characterization and mapping of the Bacillus subtilis prf gene.  
A.Reference number: I39958; MUID:87083406; PMID:3098734  
A.Accession: I39958  
A.Status: preliminary; translated from GB/EMBL/DDBJ  
A.Molecule type: DNA  
A.Residues: 1-60 <RE2>  
A.Cross-references: GB:M15318; NID:g143356; PIDN:AAA22672.1; PID:g143357  
R.Kumai, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertozzi, C.; Bron, S.; Brouillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chodura, J.; Ehrlich, S.D.; Emmertson, P.T.; Enlian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A.Authors: Foulger, D.; Fritz, C.; Fujita, M.; Funai, S.; Galizzi, A.; Gallenbach, J.; Harwood, C.R.; Hernat, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulio, M.F.; Koester, P.; Konigstein, G.; Krogh, V.; Kumano, M.; Kurita, K.; Lapilus, A.; Lardinois, A.; Lauter, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel-Rieger, M.; Rivolta, C.; Kocha, E.; Koche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanton, A.; Schlegel, S.; Schroter, R.; Scottone, F.; Sekowska, A.; Seron, A.; Teuchmann, M.; Yamakoshi, A.; Tanaka, T.; Terpilata, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wißat, A.; Yanamoto, H.; Yanane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
A.Authors: Yoshikawa, H.F.; Zundstein, E.; Yoshikawa, H.; Danchin, A.  
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A.Reference number: A69580; MUID:98044033; PMID:9384377  
A.Accession: A69614  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A.Molecule type: DNA  
A.Residues: 1-60 <KUN>  
A.Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14112.1; PID:e183641;  
A.Experimental source: strain 168  
C.Genetics:  
A.Gene: degR

Query Match 3.4%; Score 49.5; DB 2; Length 60;  
Best Local Similarity 35.9%; Pred. No. 4.1e+03;  
Matches 14; Conservative 8; Mismatches 14; Indels 3; Gaps 2;

QY 237 VDRPMEVTSKDSGIPYSRI--FTDSTAKKGKPGDSY 273  
| | : : : : | | | | |  
Db 1 MDXKD.LILHKTPIETIYSDLELAD-IAXKKSPMEXY 38

RESULT 9  
A63286  
hyprotheical protein PAZ883 [imported] - Pseudomonas aeruginosa (strain PAO1)  
C.Species: Pseudomonas aeruginosa  
C.Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C.Accession: A63286  
R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.U.; Brown, S.; Chan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.  
A.Reference number: A62950; MUID:20437337; PMID:10984043  
A.Accession: A63286  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-55 <STO>  
A.Cross-references: GB:AE004714; GB:AE004091; NID:g9948965; PIDN:AAG06271.1; GSPDB:GN001  
A.Experimental source: strain PAO1

C:Genetics: PA2893

Query Match 3.3%; Score 48; DB 2; Length 55;  
Best Local Similarity 25.4%; Pred. No. 4.6e+03;  
Matches 15; Conservative 16; Mismatches 14; Indels 14; Gaps 3;

OY 119 HAWL-----NLENGIYSKNIKAQOLAKDPKNKETYEKNIAKAYVAKLEKDKEAKR 170  
DB 4 YAMLIIVDALSGIVGILLILDRSAKPLPSEQLKRHERRN----AEIDA--QEAQR 55

RESULT 10

ACI756  
hypothetical protein lin2592 [imported] - Listeria innocua (strain Cllp11262)  
C:Species: Listeria innocua  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
R:Glaser P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fstl, H.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schuener, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: ACI756  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-57 <GL>  
A:Cross-references: GB:AL592022; PIDN:CAC97819.1; PID:G16415129; GSPDB:GN00178  
A:Experimental source: strain Cllp11262  
C:Genetics:  
A:Gene: lin2592

Query Match 3.2%; Score 47.5; DB 2; Length 57;  
Best Local Similarity 27.6%; Pred. No. 5.3e+03;  
Matches 16; Conservative 16; Mismatches 13; Indels 13; Gaps 4;

OY 122 INLENGIYY-SKNIAKOLIADPKNKETYEKNIAKAYVAKLEKDKEA--SKFPAI 174  
DB 7 INEANSILHRKSKEISKIT-KTP-----RDLEFPAIGDKLSQEWMVDYKNELAI 56

RESULT 11

B70243  
hypothetical protein BBI27 - Lyme disease spirochete plasmid I/lp28-4  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
R:Glaser P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fstl, H.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schuener, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: ACI756  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-59 <GL>  
A:Cross-references: GB:AE000789; NID:g2690079; PIDN:AAC66217.1; PID:g2690114; TIGR:BBI2  
A:Experimental source: strain B31  
C:Genetics:  
A:Genome: plasmid

Query Match 3.2%; Score 47.5; DB 2; Length 59;  
Best Local Similarity 29.3%; Pred. No. 5.5e+03;  
Matches 17; Conservative 11; Mismatches 19; Indels 11; Gaps 2;

OY 135 AKOLIA-----KDPKKETYEKNIAKAYVAKLEKDKEAKSFPDAENKCLVTSE 165  
DB 3 AKERVAAGWQVLVDYDYNKLKTIDINKLSYNDLVNESEEQATK---AENLEEPIVSK 56

## RESULT 12

B69606

Spore coat protein cofil - *Bacillus subtilis*C/Species: *Bacillus subtilis*

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C/Accession: B69606

R/Kunet, P.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chd

A/Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot

A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Yano, K.; Yata, K.; Yoshida, K

A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: B69606

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-48 &lt;KUN&gt;

A/Cross-references: GB:229113; GB:AL009126; NID:92634090; PIDN:CAB13682.1; PID:e1183457;

A/Experimental source: strain 168

C/Genetics:

A/Gene: cofil

## Query Match

Best Local Similarity 3.2%; Score 47; DB 2; Length 48;

Matches 15; Conservative 6; Mismatches 12; Indels 26; Gaps 2;

QY 35 DLHSIVPIGDPPHEPFLPDEAKTSNADVFVNGINLDEGQAFGLVNAQKTKNK 93

DB 9 DMHGNAPRGHNGQPEPL-----SG-----SKVKNRNRHTRK 41

## RESULT 13

D81277

Hypothetical protein Cj1335 [imported] - *Campylobacter jejuni* (strain NCTC 11168)C/Species: *Campylobacter jejuni*

C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C/Accession: D81277

R/Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barril

Nature 403, 665-668, 2000

A/Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp

A/Reference number: A81250; MUID:20150912; PMID:10688204

A/Accession: D81277

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-60 &lt;PAR&gt;

A/Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:CAB73762.1; PID:9696876

A/Experimental source: serotype O2, strain NCTC 11168

C/Genetics:

A/Gene: Cj1335

## Query Match

Best Local Similarity 3.2%; Score 47; DB 2; Length 60;

Matches 12; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

QY 143 PKNKETEKULKAY--VAKLEKDKKAKSKPDAL 174

DB 5 PTKKELFNKNIEALNITLKESEIKSKFEELI 38

## RESULT 14

C46743

corneal keratan sulfate proteoglycan core protein 37A - bovine (fragments)

C/Species: *Bos primigenius taurus* (cattle)

C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 03-May-1994

C/Accession: C46743

R/Funderburgh, J.L.; Funderburgh, M.L.; Brown, S.J.; Vergnes, J.P.; Hassell, J.R.; Mann

J.T. Biol. Chem. 268, 11874-11880, 1993

A/Title: Sequence and structural implications of a bovine corneal keratan sulfate prote

A/Accession: C46743

A/Status: preliminary

A/Molecule type: protein

A/Residues: 1-58 &lt;FUN&gt;

## Query Match

Best Local Similarity 48.4%; Score 46.5; DB 2; Length 58;

Matches 15; Conservative 3; Mismatches 8; Indels 5; Gaps 3;

QY 192 SKAYGPSATWINEE--EETPOISL 219

DB 19 SFSYG-PS-YIWYLYENMLIETIPDQFSNL 47

## RESULT 15

B82146

Conserved hypothetical protein VC1876 [imported] - *Vibrio cholerae* (strain N16961 serogC/Species: *Vibrio cholerae*

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C/Accession: B82146

R/Haidelberg, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.

Charadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,

L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: B82146

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 159 &lt;HET&gt;

A/Cross-references: GB:AE004263; GB:AE003852; NID:9656399; PIDN:AAF95024.1; GSPDB:GN00

A/Experimental source: serogroup O1, strain N16961; biotype El Tor

C/Genetics:

A/Gene: VC1876

A/Map position: 1

## Query Match

Best Local Similarity 3.1%; Score 46; DB 2; Length 59;

Matches 10; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

QY 144 KKKETEKULKAYVAKLEKDKKAKSKPDAL 176

DB 15 KKKETEKULKAYVAKLEKDKKAKSKPDAL 176

Search completed: February 11, 2004, 16:51:37  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:41:21 ; Search time 17 Seconds

(without alignments)  
793.921 Million cell updates/sec

Title: US-09-869-677A-2

Perfect score: 1475

Sequence: 1 SSTGAKTAXSDKLKVVATNS.....PGDSYYAMMKMNDKISEGL 287

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 5459

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.5	4.0	50	1 AEP_MESMA	P15228 haemobutinus
2	52	3.5	24	1 IR31_HAEIN	P35756 haemophilus
3	50	3.4	52	1 YF62_HAEIN	P4254 haemophilus
4	49.5	3.4	60	1 DEGR_BACSU	P0563 bacillus su
5	47	3.2	48	1 COTL_BACSU	P71032 bacillus su
6	45.5	3.1	49	1 YOR7_BESP	Q38442 bacterioph
7	45.5	3.1	53	1 MTPB_SULTO	P23039 sulfobact
8	45	3.1	38	1 RL36_BUCAP	Q83970 buchiera ap
9	45	3.1	48	1 RL19_SINAL	P15762 sinapis alb
10	44.5	3.0	39	1 SC62_CAMPA	P82009 canis fami
11	44	3.0	36	1 Y613_ARCFU	Q29636 archaeoglob
12	44	3.0	45	1 PBPO_PARSC	P33312 parabactin
13	43.5	2.9	52	1 RL33_CAMJE	Q93138 campylobact
14	43.5	2.9	58	1 RS21_CHLTR	Q84346 chlamydia t
15	43	2.9	50	1 SI0D_CHICK	P51964 gallus gall
16	43	2.9	55	1 RL33_BUCBP	Q83949 buchiera ap
17	42.5	2.9	58	1 RS21_CHLPN	Q82910 chlamydia p
18	42.5	2.9	59	1 RL33_BORBU	Q51337 borrelia bu
19	42	2.8	38	1 RL36_BUCAI	P57570 buchiera ap
20	42	2.8	46	1 OPT_BOVIN	P58874 bos taurus
21	42	2.8	58	1 NINE_BPH19	Q48424 bacterioph
22	42	2.8	58	1 R21C_ANASP	Q83940 anabaena sp
23	41	2.8	38	1 PSBW_ANASP	Q83947 anabaena sp
24	41	2.8	52	1 RL33_CHLMU	Q93137 chlamydia m
25	41	2.8	59	1 FERI_DESDN	Q93138 chlamydia m
26	41	2.8	60	1 Y151_UREPA	Q93138 chlamydia m
27	40.5	2.7	56	1 RS21_BACSU	P1478 ureaplasma
28	40.5	2.7	57	1 FERI_DESGI	Q93138 chlamydia m
29	40.5	2.7	58	1 FERI_DESGI	Q93138 chlamydia m
30	40.5	2.7	58	1 FERI_DESGI	Q93138 chlamydia m
31	40	2.7	47	1 Y820_METUA	P00209 desulfobact
32	40	2.7	49	1 R333_LACTIA	Q34102 lactococcu
33	40	2.7	53	1 RARC_BPP22	P03050 bacterioph

34	40	2.7	59	1 Y6K4_BPP22	P57017 bacterioph
35	39.5	2.7	55	1 V6K_BYDVP	P09518 barley yell
36	39	2.6	50	1 RK32_IORTA	Q99b5 lotus japon
37	39	2.6	55	1 ATP8_PELSU	Q79674 pelomedusa
38	39	2.6	57	1 SX16_MOUSE	Q62247 mus musculu
39	39	2.6	57	1 YC18_AGLNE	P48446 aglaetochami
40	39	2.6	58	1 ACEA_HELAN	P20639 helianthus
41	39	2.6	58	1 YC18_PORPU	P51366 porphyra pu
42	39	2.6	59	1 MTRC_METTE	Q93933 methanosarc
43	38.5	2.6	53	1 RUB3_CHITE	P58025 chlorobium
44	38.5	2.6	57	1 VG18_BPMU5	Q05224 mycobacteri
45	38.5	2.6	60	1 Y11A_METUA	P81303 methanococc

## ALIGNMENTS

RESULT 1	AEP_MESMA	STANDARD;	PRT;	50 AA.
AC	P15228;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Anti-epilepsy peptide (ABP) (Fragment).			
OS	Mesobutius martenesi (Manchurian scorpion) (Buthus martenesi).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;			
OC	Butioidae; Butiidae; Mesobutius.			
OX	NCBI_TaxID=34649;			
RN	[1]			
RP	SEQUENCE.			
RC	TISSUE-VELOC.			
RX	MEDLINE=89193444; PubMed=2930463;			
RA	Zhou X.-H., Yang D., Zhang J.-H., Liu C.-M., Lei K.-J.,			
RT	"Purification and N-terminal partial sequence of anti-epilepsy			
RT	peptide from venom of the scorpion Butius martenesi Karsch.";			
RL	Biochem. J. 257:509-517(1989).			
CC	- FUNCTION: Shows anti-epileptic activity.			
CC	- SUBCELLULAR LOCATION: Secreted.			
CC	- TISSUE SPECIFICITY: Expressed by the venom gland.			
CC	- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.			
CC	ALPHA-TOXIN SUBFAMILY.			
DR	PIR; S02174; S02174.			
DR	HSSP; P01494; 2SN3.			
DR	InterPro; IPR003614; Knott1.			
DR	InterPro; IPR002061; Scorpion_toxin1.			
DR	Pfam; PF00537; toxin_3; 1.			
DR	ProDom; PD000908; Scorpion_toxin1; 1.			
DR	SMART; SM00505; Knott1; 1.			
FT	NON_TER 50			
SQ	SEQUENCE 50 AA; 5428 MW; B9EB616620B87D4C CRC64;			
Query Match 4.0%; Score 58.5; DB 1; Length 50;				
Best Local Similarity 34.1%; Pred. No. 4.3e+02;				
Matches 15; Conservative 9; Mismatches 15; Indels 5; Gaps 3;				
OY	175 AENKLT--ITSGGCFKFSKAYGPSAYIWEINTEE--EGTPD 214			
Db	7 SDNCKVSCILGNBGNKNE-CRAYGASGYCMTVLAQDCRGLPD 49			
RESULT 2				
ID	IR31_HAEIN	STANDARD;	PRT;	24 AA.
AC	P35756;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Iron-regulated 31 kDa protein (Fragment).			
OS	Haemophilus influenzae.			
OC	Bacteriia; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
OX	NCBI_TaxID=727;			

```

RN [1]
RP SEQUENCE
RC STRAIN=EAGAN / 8358;
RX MEDLINE=92210482; PubMed=1556062;
RA Hartness R.E., Chong P., Klein M.H.;
RT "Identification of two iron-repressed periplasmic proteins in
  Haemophilus influenzae.";
RL J. Bacteriol. 174:2425-2430(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN IRON UPTAKE.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: By iron deprivation.
DR Pfam; PF01297; SBP_bac_9;
KW Iron; Periplasmic.
FT NON_TER
SQ SEQUENCE 24 AA; 2538 MW; 15F8C80AA7989567 CRC64;

Query Match 3.5%; Score 52; DB 1; Length 24;
Best Local Similarity 47.6%; Pred. No. 4.7e+02;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 12 KLVVATNSIIDMKAIAGD 32
Db 1 KFKVVTFTVIQDIQNVAGD 21

RESULT 3
YF62_HAEIN
ID YF62_HAEIN STANDARD; PRT; 52 AA.
AC P44254;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein HI1562.
GN HI1562
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
  Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U32831; AAC23220.1; -
DR PIR; C64036; C64036.

```

```

DR TIGR; HI1562; -
KW Complete proteome.
SQ SEQUENCE 52 AA; 6035 MW; 459726C0AD4CC18E CRC64;

Query Match 3.4%; Score 50; DB 1; Length 52;
Best Local Similarity 28.3%; Pred. No. 1.7e+03;
Matches 15; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

OY 138 LIANDPKKTEYKRLKAVAKLEKLDKRAKSPDAIKKLIYSEGCFKY 130
Db 1 LMSDKPK-----VLKKGLEKSKSKKKKTFGACD-----LRSGGCDKY 41

RESULT 4
ID DEGR_BACSU STANDARD; PRT; 60 AA.
AC P06563;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Regulatory protein deGR.
GN DEGR OR PRTR.
OS Bacillus subtilis, and
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423; 86029;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.subtilis; STRAIN=168 / Marburg;
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azavedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
  the serA and kgd loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2005-2016(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=B.subtilis; STRAIN=168 / Marburg;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azavedo V., Bertieri M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Bruch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Ertlin K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsteppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaert-Blandhard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,
RA Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,
RA Priesean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wandut R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

```

RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*."; <http://www.ncbi.nlm.nih.gov/PMOBSUM/summary.cgi?acc=U00096>  
 RL Nature 390:249-256(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=B.s.natto;  
 RX MEDLINE=8616805; PubMed=3082853;  
 RA Nagamori Y., Tanaka T.;  
 RT "Molecular cloning and nucleotide sequence of a DNA fragment from  
 RT *Bacillus natto* that enhances production of extracellular proteases  
 RT and levananucrase in *Bacillus subtilis*."; <http://www.ncbi.nlm.nih.gov/PMOBSUM/summary.cgi?acc=U00096>  
 RL J. Bacteriol. 166:20-28(1986).  
 CC -1- FUNCTION: LEADS, IN B. SUBTILIS, TO ENHANCED PRODUCTION OF  
 CC LEVANANUCLEASE, ALKALINE PROTEASE, AND NEUTRAL PROTEASE. IN B. NATTO  
 CC IT IS NONESSENTIAL FOR GROWTH OR EXPRESSION OF PROTEASES AND  
 CC LEVANANUCLEASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M15318; AAA22672.1; -;  
 DR EMBL; U77246; AAA96622.1; -;  
 DR EMBL; Z99115; CAB14112.1; -;  
 DR EMBL; M12917; AAA22671.1; -;  
 DR PIR; I39957; I39957;  
 DR Subtilist; BG10699; degn.  
 KW Transcription regulation; Complete proteome.  
 SQ SEQUENCE 60 AA; 7109 MW; C7B7F892FF0AB131 CRC64;  
 QY Query Match 3.4%; Score 49.5; DB 1; Length 60;  
 Best Local Similarity 35.9%; Pred. No. 2.1e+03;  
 Matches 14; Conservative 8; Mismatches 14; Indels 3; Gaps 2;  
 QY 237 VDRPMTVSKSGIPYSEI--FTDSIAKKKPGDSY 273  
 Db 1 MDDKDLILHKTFFIYSDLEBLD-IKKKPKSMWKY 38  
 ID COTL\_BACSU STANDARD; PRT; 48 AA.  
 AC P71032;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Spore coat protein L.  
 GN COTL.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97221587; PubMed=9068633;  
 RA Henriques A.O., Beall B.W., Moran C.P., Jr.;  
 RT "Coat of *Bacillus subtilis*, a member of the alpha-crystallin family of  
 RT stress proteins, is induced during development and participates in  
 RT spore outer coat formation"; <http://www.ncbi.nlm.nih.gov/PMOBSUM/summary.cgi?acc=U00096>  
 RL J. Bacteriol. 179:1887-1897(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Betero M.G., Besieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Cornerton I.F., Cummings N.D., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Enlrich S.D., Emmerson P.T.,  
 RA Ertian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Gallerton N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grand G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaser-Bianchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nodack M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Portwoll S., Prescott A.M.,  
 RA Prescean E., Pujic P., Punnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadle Y.,  
 RA Sato T., Scanlan E., Schleich S., Schreier R., Scottone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solito B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemura K.,  
 RA Takeuchi M., Takashi A., Tanaka T., Terstra P., Togonni A.,  
 RA Toesato V., Tschyama S., Vandenbol M., Vanner F., Vassarotti A.,  
 RA Viart A., Wambut R., Wedler E., Wedler H., Weitzneger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*."; <http://www.ncbi.nlm.nih.gov/PMOBSUM/summary.cgi?acc=U00096>  
 RL Nature 390:249-256(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U72073; AAC44988.1; -;  
 DR EMBL; Z99113; CAB13682.1; -;  
 DR PIR; B69606; B69606;  
 DR Subtilist; BG11921; cotL.  
 KW Sporulation; Complete proteome.  
 SQ SEQUENCE 48 AA; 5431 MW; 959773826D0C9217 CRC64;  
 QY Query Match 3.2%; Score 47; DB 1; Length 48;  
 Best Local Similarity 25.4%; Pred. No. 2.4e+03;  
 Matches 15; Conservative 6; Mismatches 12; Indels 26; Gaps 2;  
 QY 35 DLHSIVIGODPHREYPLPEDAKTSNADVIYFNGINLEDDGQAWFKLVKNAOKTKNK 93  
 Db 9 DMKNAKGNHFGQPEPL-----SG-----SKYKRNHTRK 41  
 ID YOR7\_BPSP STANDARD; PRT; 43 AA.  
 AC Q38442;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein in siz 5' region (ORF7) (Fragment).  
 OS Bacteriophage SPPL.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
 OC Lambda-like viruses.  
 OX NCBI\_TaxID=10724;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92260540; PubMed=1583695;  
 RA Tavares P., Santos M.A., Lutz R., Morelli G., de Lencastre H.,  
 RA Trautner T.A.;  
 RT "Identification of a gene in *Bacillus subtilis* bacteriophage SPPL  
 RT determining the amount of packaged DNA."; <http://www.ncbi.nlm.nih.gov/PMOBSUM/summary.cgi?acc=U00096>  
 RL J. Mol. Biol. 225:81-92(1992).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X56064; CMA39542.1; -  
 KW Hypothetical protein.  
 PT NON\_TER  
 SQ SEQUENCE 43 AA; 5103 MW; 45600C1B10F1691D CRC64;

Query Match  
 Best Local Similarity 3.1%; Score 45.5; DB 1; Length 43;  
 Matches 12; Conservative 7; Mismatches 14; Indels 1; Gaps 1;

Qy 142 DPKNKEYKNLKVAVLTKER-LDREAKSKFDAT 174  
 Db 3 EPONOEIDKYLDNITQAEKRLDKVPASRLKEI 36

RESULT 7  
 ID ID\_MTYPE\_SUTLO STANDARD; PRT; 59 AA.  
 AC P23039;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Membrane-associated ATPase epsilon chain (EC 3.6.3.14) (SUL-ATPase  
 DE epsilon)  
 GN ATP5E OR STS172.  
 OS Sulfolobus tokodaii, and  
 OS Sulfolobus acidocaldarius.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 CC Sulfolobus.  
 NC NCB1\_TaxID=11955, 2285;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=8.acidocaldarius;  
 RA MEDLINE=91072342; PubMed=2147683;  
 RT Denda K., Konishi J., Hajiro K., Oshima T., Date T., Yoshida M.;  
 RT "Structure of an ATPase operon of an acidothermophilic  
 RT archaeobacterium, Sulfolobus acidocaldarius.";  
 RL J. Biol. Chem. 265:21509-21513(1990).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=8.tokodaii; STRAIN=JCM 10545 / 7;  
 RA MEDLINE=21456156; PubMed=11572479;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Anai A., Kousugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguni A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kakuchi H.;  
 RT "Complete genome sequence of an aerobic thermacidophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
 CC H(+) (out)  
 CC -1- SUBUNIT: SUL-ATPASE IS COMPOSED OF SIX (OR FIVE?) SUBUNITS:  
 CC ALPHA, BETA, DELTA, GAMMA, C (PROTEOLIPID), AND POSSIBLY EPSILON.  
 CC -1- SIMILARITY: SOME, TO E.HRAE NTPH.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M57238; AAA72942.1; -

DR EMBL; A6000986; BAB66506.1; -  
 DR PIR; C36493; C36493.  
 KW Hydroxylase; Hydrogen ion transport; Complete proteome.  
 SQ SEQUENCE 59 AA; 7038 MW; 4C07944C732225D4 CRC64;

Query Match  
 Best Local Similarity 3.1%; Score 45.5; DB 1; Length 59;  
 Matches 20; Conservative 10; Mismatches 17; Indels 7; Gaps 4;

Qy 111 SEKKEDPHAWLN-LENGIISKNIAKQIAKPKNKEYKNLKVAVLTKER 163  
 Db 2 SEIDKSTIDKYNILSKLDQKKN--ELISK--INWE-YEKTLKQRLDELEKT 49

RESULT 8  
 ID ID\_RL36\_BUCAP STANDARD; PRT; 36 AA.  
 AC Q8K970;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 50S ribosomal protein L36.  
 GN RPLJ3 OR BUSG484.  
 OS Buchnera aphidicola (subsp. Schizaphis graminum).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 NC NCB1\_TaxID=98794;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22084549; PubMed=12089438;  
 RA Tamae I., Klasson L., Canback B., Naeslund A.K., Eriksson A.-S.,  
 RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;  
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";  
 RL Science 296:2376-2379(2002).  
 CC -1- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; A6014124; AAM68027.1; -  
 DR HAMAP; MF\_00251; -; 1.  
 DR InterPro; IPR000473; Ribosomal\_L36.  
 DR Pfam; PF00444; Ribosomal\_L36; 1.  
 DR TIGRFAMs; TIGR01022; rplM36; 1.  
 DR PROSITE; PS00828; RIBOSOMAL\_L36; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 38 AA; 4434 MW; 29EB845EA3E3CAF1 CRC64;

Query Match  
 Best Local Similarity 3.1%; Score 45; DB 1; Length 38;  
 Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 128 IYISKNIKQIAKQIAKPKNKEYKNLKVAVLTKER 147  
 Db 16 IYRKNVVRVICTNDPKHQ 35

RESULT 9  
 ID ID\_RL19\_SINNL STANDARD; PRT; 48 AA.  
 AC P15762;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chloroplast 30S ribosomal protein S19 (Fragment).  
 GN RPS19.  
 OS Sinapis alba (white mustard) (Brassica hirta).  
 OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Sinapis.  
 OX NCBI\_TaxID=3728;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Albatros;  
 RX MEDLINE=90192092; PubMed=2315019;  
 RA Nickelsen J., Link G.;  
 RT "Nucleotide sequence of the mustard chloroplast genes trnH and  
 trpL";  
 RL Nucleic Acids Res. 18:1051-1051(1990).  
 CC -1- FUNCTION: Protein S19 forms a complex with S13 that binds strongly  
 CC to the 16S ribosomal RNA (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X17331; CAC35517.1; -.  
 DR HSSP: P80381; 10KF.  
 DR HAMAP: MF\_00531; -; 1.  
 DR InterPro: IPR002222; Ribosomal\_S19.  
 DR Pfam: PF00203; Ribosomal\_S19; 1.  
 DR ProDom: PD001012; Ribosomal\_S19; 1.  
 DR PROSITE: PS00323; RIBOSOMAL\_S19; PARTIAL.  
 KM Ribosomal protein; rRNA-binding; Chloroplast.  
 FT INIT MET 0 BY SIMILARITY.  
 FT NON\_TER 48 48  
 SQ SEQUENCE 48 AA; 5770 MW; 2AASADA7B6B2C58B CRC64;  
 QY Query Match 3.1%; Score 45; DB 1; Length 48;  
 Best Local Similarity 34.5%; Pred. No. 3.2e+03;  
 Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;  
 OY 9 KSDKLKVVATNSIADMTKAIAGDKIDLH 37  
 17 KIEKLNTKAEKQILITWSRAITGRRELIN 45  
 RESULT 10  
 SC62\_CANFA STANDARD; PRT; 39 AA.  
 ID SC62\_CANFA  
 AC P82009;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Putative translocation protein sec62 homolog (Fragments).  
 OS Canis familiaris (Dog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=20319006; PubMed=10860986;  
 RA Tyedmers J., Lerner M., Bies C., Dudek J., Skowronek M., Haas I.,  
 RA Heim N., Naeiraiczky W., Volkmer J., Zimmermann R.;  
 RT "Homologs of the yeast Sec complex subunits Sec62p and Sec63p are  
 RT abundant proteins in dog pancreas microsomes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:7214-7219(2000).  
 CC -1- FUNCTION: REQUIRED FOR PREPROTEIN TRANSLOCATION.  
 CC -1- SUBUNIT: COMPLEX THAT CONTAINS SEC61, SEC62 AND SEC63.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -1- TISSUE SPECIFICITY: Pancreas.  
 CC Protein transport; Translocation; Transmembrane;  
 KM Endoplasmic reticulum.

FT NON\_TER 1 1  
 FT NON\_CONS 18 19  
 FT NON\_TER 39 39  
 SQ SEQUENCE 39 AA; 4497 MW; 6E77E7237BD5DF79 CRC64;  
 QY Query Match 3.0%; Score 44.5; DB 1; Length 39;  
 Best Local Similarity 29.4%; Pred. No. 2.7e+03;  
 Matches 10; Conservative 5; Mismatches 8; Indels 11; Gaps 1;  
 OY 94 DYFASVSGDIVYLV-----EGASEKGE 116  
 6 DYFIASRAVDYXLMKDYDKIKKEXKGEKSGE 39  
 Db  
 RESULT 11  
 Y619\_ARCFU STANDARD; PRT; 36 AA.  
 ID Y619\_ARCFU  
 AC O29636;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AF0619.  
 GN AF0619.  
 OS Archaeoglobus fulgidus.  
 CC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 CC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwim M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kervage A.R., Graham D.E., Kyriades N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirtness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus";  
 RL Nature 390:364-370(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001062; AAB90628.1; -.  
 DR PIR: C69327.  
 DR TIGR: AF0619; -.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 36 AA; 4285 MW; AB516CB8F2DD1C22 CRC64;  
 QY Query Match 3.0%; Score 44; DB 1; Length 36;  
 Best Local Similarity 37.5%; Pred. No. 2.6e+03;  
 Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
 OY 147 EYENKIKAVATLEKIDKEAKSK 170  
 11 EYVRKAIABYLEKLEIEIEVSGK 34  
 Db  
 RESULT 12  
 PBPO\_PARSC STANDARD; PRT; 45 AA.  
 ID PBPO\_PARSC  
 AC P83312;



DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Paratubopurin.  
 OS Paratubus schlechteri (Scorpion).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
 OC Butchidae; Butchidae; Paratubus.  
 NCBI\_Taxid=190110;  
 [1]  
 RP SEQUENCE. FUNCTION. SUBUNIT. SUBCELLULAR LOCATION. TISSUE SPECIFICITY.  
 RP AND MASS SPECTROMETRY.  
 RC TISSUE=Venom;  
 RA Verdonck F., Bosteels S., Deemert J., Moerman L.F.A., Noppe W.,  
 RA Willems J., Tytgat J., Van der Walt J.;  
 RT "A novel class of pore-forming peptides in the venom of Paratubus  
 RT schlechteri Purcell (Scorpiones: Butchidae).";  
 RT Chimbabwa 16:247-260(2000).  
 RN [2]  
 RP FUNCTION. CIRCULAR DICHROISM ANALYSIS, AND SYNTHESIS.  
 RC TISSUE=Venom;  
 RX MEDLINE=2241537; PubMed=12354111;  
 RA Moerman L.F.A., Bosteels S., Noppe W., Willems J., Clynen E.,  
 RA Schoofs L., Thevissen K., Tytgat J., Van Eldere J., van der Walt J.,  
 RA Verdonck F.;  
 RT "Antibacterial and antifungal properties of alpha-helical, cationic  
 RT peptides in the venom of scorpions from southern Africa.";  
 RT Eur. J. Biochem. 269:479-481(2002).  
 CC -1- FUNCTION: Induces a leak current in voltage-clamped dorsal root  
 CC ganglion cells of rats by forming pores. Degranulates human  
 CC granulocytes and has a weak hemolytic activity against human red  
 CC blood cells. Has antibacterial activity against Gram-negative  
 CC bacteria but is less active against Gram-positive bacteria. Has  
 CC antifungal activity.  
 CC -1- SUBUNIT: Monomer and homodimer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -1- MASS SPECTROMETRY: MW=5030.3; METHOD=MALDI.  
 KM Toxin; Neurotoxin; Antidiabetic; Fungicide; Hemolysis.  
 FT UNSURE 11 11  
 FT UNSURE 44 44  
 FT UNSURE 45 45  
 SQ SEQUENCE 45 AA; 4995 MW; 3E9F71425F81AFA CRC64;  
 Query Match 3.0%; Score 44; DB 1; Length 45;  
 Best Local Similarity 37.5%; Pred. No. 3.5e+03;  
 Matches 15; Conservative 6; Mismatches 9; Indels 10; Gaps 3;  
 Oy 77 QAWFTKLVKQNAQ---KTNNKDYFAVSGIDIVYIEGASEK 113  
 Db 9 KAWSKLAKKLRANGKEMKNDY---AKGL---LEGGSEE 41  
 RESULT 13  
 RL33 CAMJE STANDARD; PRT; 52 AA.  
 AC OPR18;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 50S ribosomal protein L33.  
 GN RPL33 OR C10471.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 NCBI\_Taxid=197;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 RA Jagers K., Kariyeh A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,

RA Whitehead S., Barrell B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 CC -1- SIMILARITY: BELONGS TO THE L33P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL139075; CAB75109.1;  
 DR PIR; B81392; B81392.  
 DR HAMAP; MF 00294; -- 1.  
 DR InterPro; IPR001705; Ribosomal\_L33.  
 DR Pfam; PF00471; Ribosomal\_L33; 1.  
 DR ProDom; PD002595; Ribosomal\_L33; 1.  
 DR TIGRFAMs; TIGR01023; rpl33.bact; 1.  
 DR PROSITE; PS00582; RIBOSOMAL\_L33; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 52 AA; 6156 MW; 2502CA2506D3E5FE CRC64;  
 Query Match 2.9%; Score 43.5; DB 1; Length 52;  
 Best Local Similarity 48.3%; Pred. No. 4.5e+03;  
 Matches 14; Conservative 3; Mismatches 9; Indels 3; Gaps 2;  
 Oy 145 NKETVKK-NIKVAVALKK--LDKSKSK 170  
 Db 23 SKATTEKLEKCYPRKXHTLKEVYK 51  
 RESULT 14  
 RS21 CHLTR STANDARD; PRT; 58 AA.  
 AC 0843T6;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S21.  
 GN RPSU OR RS21 OR CP342 OR TC0620.  
 OS Chlamydia trachomatis, and  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_Taxid=813, 83560;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C. trachomatis; STRAIN=D/TW-3/Cx;  
 RX MEDLINE=9900809; PubMed=9784136;  
 RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis.";  
 RL Science 282:754-759(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C. muridarum; STRAIN=Mopn / N199;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uitterlinden T., Berry K.,  
 RA Bagg S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,  
 RA Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G.,  
 RA Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39.";  
 RT Nucleic Acids Res. 28:1397-1406(2000).  
 CC -1- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

DR EMBL; AE001307; AAC67937.1; -  
 DR EMBL; AE002330; AAF39451.1; -  
 DR PIR; A71527; A71527.  
 DR PIR; E81683; E81683.  
 DR TIGR; TC0620; -  
 DR HAMAP; MF\_00358; -; 1.  
 DR InterPro; IPR001911; Ribosomal\_S21.  
 DR Pfam; PF01165; Ribosomal\_S21; 1.  
 DR PRINTS; PR00976; RIBOSOMALS21.  
 DR PRODOM; PD005521; Ribosomal\_S21; 1.  
 DR TIGRFAMs; TIGR00030; S21P; 1.  
 DR PROSITE; PS01181; RIBOSOMAL\_S21; FALSE\_NEG.  
 KW Ribosomal protein; Complete proteome.  
 SO SEQUENCE 58 AA; 6666 MW; 99AD89D04625185F CRC64;

Query Match 2.9%; Score 43.5; DB 1; Length 58;  
 Best Local Similarity 33.3%; Pred. No. 5.1e+03;  
 Matches 14; Conservative 4; Mismatches 15; Indels 9; Gaps 1;

Qy 129 IYSNIAKOLIAKDKKETEYKXLYAVAKLEKDKKAKSK 170  
 Db 18 ILKKKIDKEGILKTSKSHRYDK-----PSVKKRAKSK 50

RESULT 15  
 S10D CHICK  
 ID \_S10D CHICK STANDARD; PRT; 50 AA.  
 AC PS1926;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Vitamin D-dependent calcium-binding protein, intestinal (CABP)  
 DE (calbindin D9K) (Fragment).  
 GN CALB3 OR S100D.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95269726; PubMed=7750504;  
 RA Zanelli S.B., Boland R.L., Norman A.W.;  
 RT "cDNA sequence identity of a vitamin D-dependent calcium-binding  
 RT protein in the chick to calbindin D-9K."  
 RL Endocrinology 136:2784-2787(1995).  
 CC -1- SIMILARITY: BELONGS TO THE S-100 FAMILY.  
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

DR EMBL; S78183; AAD14276.1; -  
 DR PIR; I51287; I51287.  
 DR HSSP; P02633; 4ICB.  
 DR InterPro; IPR001751; CABP\_S100.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; ehand; 1.  
 DR Pfam; PF01023; S\_100; 1.  
 DR PRODOM; PD003407; CABP\_S100; 1.  
 DR PRODOM; PD000012; EF-hand; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.

DR PROSITE; PS00303; S100 CABP; PARTIAL.  
 KW Calcium-binding; Vitamin D.  
 FT NOW TER 1 1  
 FT CA\_BIND <1 10 EF-HAND 1 (LOW AFFINITY).  
 FT CA\_BIND 37 48 EF-HAND 2 (HIGH AFFINITY).  
 FT NOW TER 50 50  
 SO SEQUENCE 50 AA; 5646 MW; 5D4484AC31C4ED28 CRC64;

Query Match 2.9%; Score 43; DB 1; Length 50;  
 Best Local Similarity 32.0%; Pred. No. 4.6e+03;  
 Matches 16; Conservative 7; Mismatches 21; Indels 6; Gaps 2;

Qy 211 GTPDQSLIEKLVKVI----KPSALFVSSVDRRPMETVSKDSGIPYSE 256  
 Db 1 GDFVQLSK--BELKLLIQSEFPLKASSTLDNLFELDNGDGEVSYER 48

Search completed: February 11, 2004, 16:50:24  
 Job time : 32 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:40:29 ; Search time 40 Seconds  
(without alignments)  
1138.863 Million cell updates/sec

Title: US-09-869-677A-2

Perfect score: 1475  
Sequence: 1 SSTGAKTKASDKLKVATNS.....FGDSYAMMKVNLKISEGL 287

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 605656

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	8.1	27	20	AAI29825
2	117	7.9	22	21	AAI29825
3	94	6.4	20	21	AAI29825
4	89	6.0	18	21	AAI29825
5	78	5.3	17	21	AAI29825
6	72	4.9	17	21	AAI29825
7	62	4.2	16	21	AAI29825
8	62	4.2	22	22	AAI29825
9	62	4.2	22	22	AAI29825

10	61	4.1	60	22	AAI29825	Proionibacterium
11	60	4.1	60	22	AAI29825	Arabidopsis thalia
12	60	4.1	60	22	AAI29825	Arabidopsis thalia
13	60	4.1	60	22	AAI29825	Basic leucine zipp
14	58	3.9	58	21	AAI29825	Human colon cancer
15	57.5	3.9	59	21	AAI29825	Arabidopsis thalia
16	56	3.8	18	21	AAI29825	ELISA control pept
17	55	3.7	38	17	AAI29825	Cenorchabitis y06
18	54	3.7	51	22	AAI29825	Propionibacterium
19	54	3.7	60	22	AAI29825	Arabidopsis thalia
20	54	3.7	60	22	AAI29825	Arabidopsis thalia
21	54	3.7	60	22	AAI29825	Basic leucine zipp
22	53.5	3.6	56	22	AAI29825	Chlamydia trachoma
23	53.5	3.6	59	22	AAI29825	Human cardiovascular
24	53.5	3.6	60	16	AAI29825	Meg albumin bindin
25	53	3.6	60	22	AAI29825	Arabidopsis thalia
26	53	3.6	60	22	AAI29825	Arabidopsis thalia
27	53	3.6	60	22	AAI29825	Basic leucine zipp
28	51.5	3.5	37	22	AAI29825	Human bone marrow
29	51.5	3.5	37	22	AAI29825	Human bone marrow
30	51.5	3.5	37	23	AAI29825	Human peptide enco
31	51.5	3.5	35	21	AAI29825	Zea mays protein I
32	51.5	3.5	59	22	AAI29825	Human ovarian and/
33	51.5	3.5	59	22	AAI29825	Human reproductive
34	51	3.5	22	16	AAI29825	C. albicans enolase
35	50	3.4	47	22	AAI29825	Human polypeptide
36	50	3.4	51	22	AAI29825	Human liver peptid
37	50	3.4	51	22	AAI29825	Peptide #10946 enc
38	50	3.4	51	22	AAI29825	Human brain expres
39	50	3.4	51	22	AAI29825	Human bone marrow
40	50	3.4	51	22	AAI29825	Peptide #7558 enco
41	50	3.4	55	23	AAI29825	Human ORFX protein
42	49.5	3.4	59	21	AAI29825	Helicobacter pylori
43	49.5	3.4	59	23	AAI29825	Helicobacter pylori
44	49.5	3.4	60	8	AAI29825	Bacillus exoprotea
45	49	3.3	16	18	AAI29825	Plasmodium falcipa

#### ALIGNMENTS

RESULT 1  
ID AAY29825 standard; peptide: 27 AA.  
AAY29825;  
AC AAY29825;  
DT 15-NOV-1999 (first entry)  
XX 57 kDa protein (1-30) SP37.  
DE T-cell epitope; meningococcal; pneumococcal; bactericidal;  
KW multi-oligosaccharide glycoconjugate bacterial meningitis vaccine;  
RW pneumolysin; immunogenic; immune response; vaccination; meningitis;  
KW tumour-related antigen; diagnosis; detection; anti-cancer.  
XX Synthetic.  
OS  
XX  
XX WO9942130-A1.  
XX  
XX 26-AUG-1999.  
XX  
XX 23-FEB-1999; 99WO-CA00157.  
XX  
XX 23-FEB-1998; 98US-0027956.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Chong P, Klein MH, Lindberg A;  
XX WPI; 1999-540273/45.  
XX  
XX Multivalent immunogenic molecule comprising carrier with T cell

PT epitope and many carbohydrate fragments with B cell epitopes,  
 XX particularly for vaccination against meningitis and diagnosis  
 PS Example 6; Page 53; 83pp; English.

CC The present invention describes a multivalent immunogenic molecule (I)  
 CC comprising: (i) carrier (Ia) having at least one functional T-cell  
 CC epitope; and (ii) many different carbohydrate fragments (Ib), all linked  
 CC to (Ia) and each having at least one functional B-cell epitope. (Ia)  
 CC increases the immunogenicity of (Ib). AAY29819 to AAY29826 represent  
 CC potential T-cell epitopes from meningococcal and pneumococcal proteins,  
 CC used in the exemplification of the present invention. (I) are used to  
 CC generate an immune response, specifically for protective vaccination  
 CC against meningitis (Streptococcus pneumoniae or Neisseria meningitidis),  
 CC but also against tumour-related antigens and antigens from other  
 CC bacteria, e.g. Escherichia coli, Salmonella typhi, Streptococcus mutans,  
 CC Streptococcus neoformans, Klebsiella, Staphylococcus aureus or  
 CC Pseudomonas aeruginosa, to detect, by complex formation, (I)-reactive  
 CC antibodies and to raise (Ib)-specific antibodies, either for diagnostic  
 CC detection of the corresponding antigen in usual immunoassays or, if  
 CC directed against tumour antigens, for conjugation to anticancer agents.  
 CC The combination of T- and B-cell epitopes in a single vaccine results  
 CC in a strong and long-lasting humoral immunity.

SQ Sequence 27 AA;

Query Match 8.1%; Score 120; DB 26; Length 27;  
 Best Local Similarity 88.5%; Pred. No. 0.0031; Mismatches 0;  
 Matches 23; Conservative 2; Indels 0; Gaps 0;

QY 127 GIYYSNIAKQIAKDPKREYKN 152

Db 1 GIYAKNIAKQIAKDPKKNDFEYKN 26

RESULT 2  
 AAB07437  
 ID AAB07437 standard; Peptide; 22 AA.

AC AAB07437;

DT 20-OCT-2000 (first entry)

DE Antigenic peptide from a lipoprotein of an ABC transporter protein.

KW Lipoprotein; ATP-binding cassette transporter; MtsA; immune response;

OS Streptococcus pyogenes.

FT Key Location/Qualifiers

FT Misc-difference 1 /note= "this residue is not part of the native sequence"

PN WO200040729-A1.

PD 13-JUL-2000.

PF 30-DEC-1999; 99WO-GB04445.

PR 31-DEC-1998; 98GB-0028880.

PA (ACTI-) ACTINOVA LTD.

PI Bjoerck L, Janulczyk R;

DR WPI; 2000-465989/40.

XX New polypeptide comprising an ATP-binding cassette transporter of  
 PT Streptococcus pyogenes for use as a vaccine against Streptococcus and  
 PT for assays that detect immune reactivity to the polypeptides in animals  
 XX and humans -

PS Example 7; Page 31; 55pp; English.

CC AAB07437-41 represent antigenic peptides derived from a Streptococcus  
 CC pyogenes polypeptide which is lipoprotein of an ATP-binding cassette  
 CC transporter (MtsA). The polypeptide generates a protective immune  
 CC response to Streptococcus, preferably group A, such as S. pyogenes.  
 CC It is used to prepare a vaccine against Streptococcus. The new  
 CC polypeptides may be used in serological or cell mediated immune  
 CC assays for the detection of immune reactivity to the polypeptides in  
 CC animals and humans. Antibodies to MtsA protein can be detected using  
 CC an immunoassay with the polypeptides.

SQ Sequence 22 AA;

Query Match 7.9%; Score 117; DB 21; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.0041; Mismatches 0;  
 Matches 21; Conservative 0; Indels 0; Gaps 0;

QY 259 TDSIAKKKRGDSYAMMKNN 279

Db 2 TDSIAKKKRGDSYAMMKNN 22

RESULT 3

ID AAB07440 standard; Peptide; 20 AA.

AC AAB07440;

DT 20-OCT-2000 (first entry)

DE Antigenic peptide from a lipoprotein of an ABC transporter protein.

KW Lipoprotein; ATP-binding cassette transporter; MtsA; immune response;

OS Streptococcus pyogenes.

FT Key Location/Qualifiers

FT Misc-difference 1 /note= "this residue is not part of the native sequence"

PN WO200040729-A1.

PD 13-JUL-2000.

PF 30-DEC-1999; 99WO-GB04445.

PR 31-DEC-1998; 98GB-0028880.

PA (ACTI-) ACTINOVA LTD.

PI Bjoerck L, Janulczyk R;

DR WPI; 2000-465989/40.

XX New polypeptide comprising an ATP-binding cassette transporter of  
 PT Streptococcus pyogenes for use as a vaccine against Streptococcus and  
 PT for assays that detect immune reactivity to the polypeptides in animals  
 XX and humans -

Example 7; Page 31; 55pp; English.

CC AAB07437-41 represent antigenic peptides derived from a Streptococcus  
 CC pyogenes polypeptide which is lipoprotein of an ATP-binding cassette  
 CC transporter (MtsA). The polypeptide generates a protective immune  
 CC response to Streptococcus, preferably group A, such as S. pyogenes.  
 CC It is used to prepare a vaccine against Streptococcus. The new  
 CC polypeptides may be used in serological or cell mediated immune  
 CC assays for the detection of immune reactivity to the polypeptides in  
 CC animals and humans. Antibodies to MtsA protein can be detected using  
 CC an immunoassay with the polypeptides.

SQ Sequence 20 AA;

Query Match 6.4%; Score 94; DB 21; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.3; Mismatches 0; Indels 0; Gaps 0;

QY 204 EINTEEGTPDQISSLIEK 222

DB 2 EINTEEGTPDQISSLIEK 20

RESULT 4

AAB07439

ID AAB07439 standard; Peptide; 18 AA.

XX AAB07439;

DT 20-OCT-2000 (first entry)

DE Antigenic peptide from a lipoprotein of an ABC transporter protein.

KM Lipoprotein; ATP-binding cassette transporter; MtsA; immune response;

XX Streptococcus pyogenes.

FH Key Location/Qualifiers

FT Misc-difference 1 /note= "this residue is not part of the native sequence"

PN W0200040729-A1.

PD 13-JUL-2000.

PF 30-DEC-1999; 99WC-GB04445.

PR 31-DEC-1998; 98GB-0028880.

XX (ACT1-) ACTINOVIA LTD.

PI Bjorck L, Janulczyk R;

DR WPI; 2000-465989/40.

PT New polypeptide comprising an ATP-binding cassette transporter of  
 Streptococcus pyogenes for use as a vaccine against Streptococcus and  
 for assays that detect immune reactivity to the polypeptides in animals  
 and humans -

PS Example 7; Page 31; 55pp; English.

CC AAB07437-41 represent antigenic peptides derived from a Streptococcus  
 CC pyogenes polypeptide which is lipoprotein of an ATP-binding cassette  
 CC transporter (MtsA). The polypeptide generates a protective immune  
 CC response to Streptococcus, preferably group A, such as S. pyogenes.  
 CC It is used to prepare a vaccine against Streptococcus. The new  
 CC polypeptides may be used in serological or cell mediated immune  
 CC assays for the detection of immune reactivity to the polypeptides in  
 CC animals and humans. Antibodies to MtsA protein can be detected using  
 CC an immunoassay with the polypeptides.

SQ Sequence 18 AA;

Query Match 6.0%; Score 89; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 0.69; Mismatches 0; Indels 0; Gaps 0;

QY 136 KQIIADPKPKKTEYENK 152

DB 2 KQIIADPKPKKTEYENK 18

RESULT 5

AAB07438

ID AAB07438 standard; Peptide; 17 AA.

XX AAB07438;

DT 20-OCT-2000 (first entry)

DE Antigenic peptide from a lipoprotein of an ABC transporter protein.

KM Lipoprotein; ATP-binding cassette transporter; MtsA; immune response;

XX Streptococcus pyogenes.

FH Key Location/Qualifiers

FT Misc-difference 1 /note= "this residue is not part of the native sequence"

PN W0200040729-A1.

PD 13-JUL-2000.

PF 30-DEC-1999; 99WC-GB04445.

PR 31-DEC-1998; 98GB-0028880.

XX (ACT1-) ACTINOVIA LTD.

PI Bjorck L, Janulczyk R;

DR WPI; 2000-465989/40.

PT New polypeptide comprising an ATP-binding cassette transporter of  
 Streptococcus pyogenes for use as a vaccine against Streptococcus and  
 for assays that detect immune reactivity to the polypeptides in animals  
 and humans -

PS Example 7; Page 31; 55pp; English.

CC AAB07437-41 represent antigenic peptides derived from a Streptococcus  
 CC pyogenes polypeptide which is lipoprotein of an ATP-binding cassette  
 CC transporter (MtsA). The polypeptide generates a protective immune  
 CC response to Streptococcus, preferably group A, such as S. pyogenes.  
 CC It is used to prepare a vaccine against Streptococcus. The new  
 CC polypeptides may be used in serological or cell mediated immune  
 CC assays for the detection of immune reactivity to the polypeptides in  
 CC animals and humans. Antibodies to MtsA protein can be detected using  
 CC an immunoassay with the polypeptides.

SQ Sequence 17 AA;

Query Match 5.3%; Score 78; DB 21; Length 17;

Best Local Similarity 100.0%; Pred. No. 5.3; Mismatches 0; Indels 0; Gaps 0;

QY 234 ESSVDRRPMETYSKDS 249

DB 2 ESSVDRRPMETYSKDS 17

RESULT 6

AAB07441

ID AAB07441 standard; Peptide; 13 AA.

XX AAB07441;

DT 20-OCT-2000 (first entry)

DE Antigenic peptide from a lipoprotein of an ABC transporter protein.

KM Lipoprotein; ATP-binding cassette transporter; MtsA; immune response;

XX vaccine.

```

OS Streptococcus pyogenes.
XX
XX Key Location/Qualifiers
FT MISC-difference 1 /note="this residue is not part of the native sequence"
XX
XX
XX WO200040729-A1.
XX
XX 13-JUL-2000.
XX
XX 30-DEC-1999; 99WO-GB04445.
XX
XX 31-DEC-1998; 98GB-0028880.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjoerck L, Janulczyk R;
XX
XX WPI; 2000-465989/40.
XX
XX New polypeptide comprising an ATP-binding cassette transporter of
XX Streptococcus pyogenes for use as a vaccine against Streptococcus and
XX for assays that detect immune reactivity to the polypeptides in animals
XX and humans -
XX
XX Example 7; Page 31; 55pp; English.
XX
XX AA807437-41 represent antigenic peptides derived from a Streptococcus
XX pyogenes polypeptide which is lipoprotein of an ATP-binding cassette
XX transporter (Mtsa). The polypeptide generates a protective immune
XX response to Streptococcus, preferably group A, such as S. pyogenes.
XX It is used to prepare a vaccine against Streptococcus. The new
XX polypeptides may be used in serological or cell mediated immune
XX assays for the detection of immune reactivity to the polypeptides in
XX animals and humans. Antibodies to Mtsa protein can be detected using
XX an immunoassay with the polypeptides.
XX
XX Sequence 13 AA;
XX
XX Query Match 4.9%; Score 72; DB 21; Length 13;
XX Best Local Similarity 100.0%; Pred. NO. 12;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 44 QDPHEYEPLPED 55
XX
XX Db 1 QDPHEYEPLPED 12
XX
XX RESULT 7
XX ID AAY82573 standard; peptide; 56 AA.
XX
XX AAY82573;
XX
XX 28-JUL-2000 (first entry)
XX
XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.
XX
XX Copolymer; molecular weight marker; TV-marker; immune disease;
XX glatiramer acetate; autoimmune disease; antiinflammatory;
XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;
XX antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
XX antihaemetic; immunosuppressive; demyelinating disease; osteoarthritis;
XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;
XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;
XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
XX pemphigus vulgaris; systemic lupus erythematosus.
XX
XX Unidentified.
XX
XX OS
XX
XX MO200018794-A1.
XX

```

```

PD 06-APR-2000.
XX
XX 24-SEP-1999; 99WO-US22402.
XX
XX 25-SEP-1998; 98US-0101693.
XX
XX (VEDA ) VEDA RES & DEV CO LTD.
XX
XX (TEVA-) TEVA PHARM USA INC.
XX
XX Gad A, Lis D;
XX
XX WPI; 2000-317499/27.
XX
XX Copolymer 1 related polypeptides used as molecular weight markers for
XX glatiramer acetate and for treatment and prevention of immune diseases
XX
XX Claim 10; Page 14; 72pp; English.
XX
XX AA82571 to AAY82577 represent specifically claimed copolymer molecular
XX weight TV-marker polypeptides from the present invention. The present
XX invention describes polypeptides (I) for determining the molecular
XX weight of a copolymer (CP), which has an identified molecular weight
XX and an amino acid composition corresponding to the copolymer. The
XX polypeptides of the invention are used as molecular weight markers for
XX glatiramer acetate related tetrapolymers. The polypeptides may also be
XX used for treating and preventing immune diseases in a mammal. Autoimmune
XX diseases which may be treated include either cell-mediated or
XX antibody-mediated diseases. Such diseases include arthritic conditions,
XX demyelinating diseases and inflammatory conditions, e.g. multiple
XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
XX anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
XX uveoretinitis, Crohn's disease, chronic immune thrombocytopenia
XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
XX lupus erythematosus. Mediated-mediated diseases which can be treated
XX include host-versus-graft disease, graft-versus-host disease, and
XX delayed-type hypersensitivity. The polypeptides of the invention have
XX defined molecular weights and physical properties which are analogous to
XX glatiramer acetate molecules, which makes them ideal for use as
XX molecular weight markers.
XX
XX Sequence 56 AA;
XX
XX Query Match 4.2%; Score 62; DB 21; Length 56;
XX Best Local Similarity 38.9%; Pred. NO. 5.8e+02;
XX Matches 21; Conservative 7; Mismatches 22; Indels 4; Gaps 2;
XX
XX 149 YEKNLKAYVAKLEKDEKSKFDAIENKKUIVTSECGFK-YPSKAGVGSAY 201
XX
XX Db 4 YAKKKAIVAKKKAENKAKAENKAYRAAEK--KAEKRYAEAKKAERAY 54
XX
XX RESULT 8
XX ID AAB64845 standard; Protein; 56 AA.
XX
XX AAB64845;
XX
XX 23-MAR-2001 (first entry)
XX
XX Gene 26 human secreted protein homologous amino acid sequence #131.
XX
XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
XX antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; vitrucide; fungicide;
XX ophthalmological; autoimmune disease; hyperproliferative disorder;
XX cardiovascular disorder; cerebrovascular disorder; infection; chemotaxis;
XX nervous system disorder; ocular disorder; skin aging; wound healing;
XX food additive; tissue regeneration.
XX
XX Rattus norvegicus.
XX
XX OS
XX

```

PN W0200077256-A1.  
 XX  
 XX 21-DEC-2000.  
 XX  
 XX 01-JUN-2000; 2000WO-US14963.  
 XX  
 XX 11-JUN-1999; 99US-0138631.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI; 2001-032315/04.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 XX Disclosure; Page 53; 506pp; English.  
 XX  
 CC Polynucleotide sequences AAF33095 - AAF33142 encode human secreted  
 CC proteins AAB64773 - AAB64820. Fragments of the secreted proteins and  
 CC amino acid sequences which share homology with the fragments are  
 CC represented in AAB64821 - AAB64880. The genes and proteins have  
 CC activities dependent on the tissues and cells in which they are  
 CC expressed. Examples of their activities and the activities of their  
 CC agonists and antagonists include; immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; and optalmalological. The secreted proteins, polynucleotides,  
 CC antagonists and agonists may be useful in treating, preventing and  
 CC diagnosing diseases and disorders such as autoimmune diseases e.g.  
 CC rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the  
 CC breast or liver, cardiovascular disorders e.g. cardiac arrest,  
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous  
 CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,  
 CC viruses and fungi and ocular disorders e.g. corneal infection. The  
 CC polypeptides can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues and in chemotaxis. The polypeptides can also be  
 CC used as a food additive or preservative to increase or decrease storage  
 CC capabilities. Included in the invention are sequences AAB64772 and  
 CC AAF33095 - AAF33142 which are used in the isolation and characterisation  
 CC of the nucleotide and protein sequences of the invention.  
 XX  
 SQ Sequence 56 AA;  
 Query Match 4.2%; Score 62; DB 22; Length 56;  
 Best Local Similarity 31.4%; Pred. No. 5.8e+02;  
 Matches 16; Conservative 12; Mismatches 19; Indels 4; Gaps 1;  
 Oy 130 YSKNIAQOLAKDPKXKTEYKNIKAYVAKLEKDKAKSKFPAIAENKKL 180  
 Db 4 FIKWAKKKLAKQKMSKNKRENAHLAAMLRLQEKDKHA---EVRNKKEL 50  
 RESULT 9  
 AAB64846  
 ID AAB64846 standard; Protein; 56 AA.  
 XX  
 XX AAB64846;  
 XX  
 DT 23-MAR-2001 (first entry)  
 XX  
 DE Human secreted protein sequence encoded by gene 26 SEQ ID NO:132.  
 XX  
 XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
 XX antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 XX neurotropic; neuroprotective; antibacterial; virucide; fungicide;  
 XX ophthalmological; autoimmune disease; hyperproliferative disorder;  
 XX cardiovascular disorder; cerebrovascular disorder; infection; chemotaxis;  
 XX nervous system disorder; ocular disorder; skin aging; wound healing;  
 XX food additive; tissue regeneration.

XX  
 OS Homo sapiens.  
 XX  
 XX W0200077256-A1.  
 XX  
 XX 21-DEC-2000.  
 XX  
 XX 01-JUN-2000; 2000WO-US14963.  
 XX  
 XX 11-JUN-1999; 99US-0138631.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI; 2001-032315/04.  
 XX  
 DR N-PSDB; AAF33120.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition -  
 XX  
 XX Disclosure; Page 53; 506pp; English.  
 XX  
 CC Polynucleotide sequences AAF33095 - AAF33142 encode human secreted  
 CC proteins AAB64773 - AAB64820. Fragments of the secreted proteins and  
 CC amino acid sequences which share homology with the fragments are  
 CC represented in AAB64821 - AAB64880. The genes and proteins have  
 CC activities dependent on the tissues and cells in which they are  
 CC expressed. Examples of their activities and the activities of their  
 CC agonists and antagonists include; immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; and optalmalological. The secreted proteins, polynucleotides,  
 CC antagonists and agonists may be useful in treating, preventing and  
 CC diagnosing diseases and disorders such as autoimmune diseases e.g.  
 CC rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the  
 CC breast or liver, cardiovascular disorders e.g. cardiac arrest,  
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous  
 CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,  
 CC viruses and fungi and ocular disorders e.g. corneal infection. The  
 CC polypeptides can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues and in chemotaxis. The polypeptides can also be  
 CC used as a food additive or preservative to increase or decrease storage  
 CC capabilities. Included in the invention are sequences AAB64772 and  
 CC AAF33095 - AAF33142 which are used in the isolation and characterisation  
 CC of the nucleotide and protein sequences of the invention.  
 XX  
 SQ Sequence 56 AA;  
 Query Match 4.2%; Score 62; DB 22; Length 56;  
 Best Local Similarity 31.4%; Pred. No. 5.8e+02;  
 Matches 16; Conservative 12; Mismatches 19; Indels 4; Gaps 1;  
 Oy 130 YSKNIAQOLAKDPKXKTEYKNIKAYVAKLEKDKAKSKFPAIAENKKL 180  
 Db 4 FIKWAKKKLAKQKMSKNKRENAHLAAMLRLQEKDKHA---EVRNKKEL 50  
 RESULT 10  
 AAU60040  
 ID AAU60040 standard; Protein; 60 AA.  
 XX  
 XX AAU60040;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #20936.  
 XX  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.  
 OS Propionibacterium acnes.  
 XX WO200181581-A2.  
 PN 01-NOV-2001.  
 PD 20-APR-2001; 2001WO-US12865.  
 PF 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX (CORI-) CORIXA CORP.  
 XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 PI WPI, 2001-616774/71.  
 DR N-PSDB; AAS59607.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 21235; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis), acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis,  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 60 AA;  
 Query Match 4.1%; Score 61; DB 22; Length 60;  
 Best Local Similarity 29.2%; Pred. No. 7.7e+02;  
 Matches 14; Conservative 10; Mismatches 24; Indels 0; Gaps 0;  
 QY 189 KYFSKAYGVPSAYIWEITEEGTPDQISLIEKLKIKPSALFVSS 236  
 Db 9 KMSWALSTPSAAVAVLSTERTSTPSCATRHSLTLSPARPFSSA 56  
 RESULT 11  
 AAE07977  
 ID AAE07977 standard; Protein: 60 AA.  
 AC AAE07977;  
 XX  
 XX 01-NOV-2001 (first entry)  
 DT  
 XX Arabidopsis thaliana ABF3 protein fragment containing a bZIP region.  
 DE  
 XX Abscisic acid responsive element; ABRE; ABRE-binding factor; ABF;  
 KW stress treatment; transgenic plant; environmental stress; bZIP;  
 KM basic leucine zipper.

XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 7..57  
 FT /label= bZIP\_region  
 PN US6232461-B1.  
 XX  
 PD 15-MAY-2001.  
 PF 19-SEP-2000; 2000US-0665309.  
 XX  
 PR 12-OCT-1999; 99US-0416050.  
 XX  
 PA (KORU-) KOREA KUMHO PETROCHEMICAL CO LTD.  
 PI Kim SY;  
 PI WPI, 2001-366358/38.  
 DR  
 XX  
 PT New nucleic acid encoding abscisic acid responsive element binding  
 PT factor 4, useful for generating transgenic plants, which are tolerant  
 PT to multiple environmental stresses -  
 XX  
 PS Disclosure; Fig 8A; 42pp; English.  
 XX  
 CC The invention relates to nucleic acid encoding the Abscisic acid  
 CC responsive element (ABRE)-binding factor 4 (ABF4). ABF4 belongs to  
 CC the ABF family of factors which bind abscisic acid responsive elements  
 CC in plants. Expression of ABFs is inducible by abscisic acid and various  
 CC stress treatments. ABFs have the potential to activate a large number  
 CC of abscisic acid/stress responsive genes and thus a nucleic acid  
 CC molecule encoding ABF4 is useful for generating transgenic plants  
 CC that are tolerant to multiple environmental stresses. The present  
 CC sequence is Arabidopsis thaliana ABF3 protein fragment containing a  
 CC bZIP region.  
 CC  
 SQ Sequence 60 AA;  
 Query Match 4.1%; Score 60; DB 22; Length 60;  
 Best Local Similarity 29.5%; Pred. No. 9.3e+02;  
 Matches 13; Conservative 13; Mismatches 18; Indels 0; Gaps 0;  
 QY 136 KQIIMKDPKNTETKNIKAVANLEKLDREAKSKFDATAEKK 179  
 Db 16 RESAARSRRKQAYTMEIEMAIQLEKELNELQKKQVEIMEKK 59  
 RESULT 12  
 AAE04628  
 ID AAE04628 standard; peptide: 60 AA.  
 AC AAE04628;  
 XX  
 XX 04-SEP-2001 (first entry)  
 DT  
 XX Arabidopsis thaliana ABF3 protein fragment containing a bZIP region.  
 DE  
 XX Abscisic acid responsive element binding factor 3; ABF3; ABRE; bZIP;  
 KW bZIP class transcription factor; transgenic plant; basic leucine zipper.  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 7..57  
 FT /label= bZIP\_region  
 PN US6245905-B1.  
 XX  
 PD 12-JUN-2001.  
 XX  
 PF 14-SEP-2000; 2000US-0661569.



```
XX 12-OCT-1999; 99US-0416050.
PR (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.
PA
XX Kim SY;
XX WPI; 2001-380516/40.
XX
XX New abscisic acid responsive element binding factor 2 useful for
PT activating abscisic acid/stress responsive genes, and generating
PT transgenic plants that are tolerant to multiple environmental stresses
XX
XX Disclosure; Fig 8A; 42pp; English.
XX
XX The patent discloses novel abscisic acid (ABA) responsive element
CC (ABRE) binding factors (ABFs). ABFs are basic leucine zipper (bZIP)
CC class transcription factors that bind to both G-box-like ABREs (G/ABREs)
CC and coupling element like ABREs (C/ABREs). ABFs have the potential to
CC activate a large number of abscisic acid/stress responsive genes, and
CC are therefore used to generate transgenic plants that are tolerant to
CC multiple environmental stresses.
CC The present sequence is Arabidopsis thaliana ABF3 protein fragment
CC containing a bZIP region.
XX
XX Sequence 60 AA;
SQ
XX
XX Query Match 4.1%; Score 60; DB 22; Length 60;
Best Local Similarity 29.5%; Pred. No. 9.3e+02;
Matches 13; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
QY 136 KQLIADPKNKETYEKNLKAIVAKLEKDKAKSPDAIAENKK 179
Db 16 RESAARSRARKQAVTMELEAIEIQLKELNEELQKKQVEIMEKK 59
XX
XX RESULT 13
XX AAU00281
XX ID AAU00281 standard; Protein; 60 AA.
XX
XX AAU00281;
XX
XX 25-MAY-2001 (first entry)
XX
XX Basic leucine zipper (bZIP) region of ABF3.
XX
XX Abscisic acid responsive element binding factor 3; ABF3; plant hormone;
XX drought; ABA-responsive element; ABRE; ABRE-binding factor; ABF;
XX basic leucine zipper; bZIP; G/ABRE; C/ABRE; transgenic plant;
XX environmental stress.
XX
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
XX FH Region 7..26
XX FT /note= "Basic region"
XX FT 30..57
XX FT Region /note= "Leucine zipper region"
XX
XX US6194559-B1.
XX
XX 27-FEB-2001.
XX
XX 12-OCT-1999; 99US-0416050.
XX
XX 12-OCT-1999; 99US-0416050.
XX
XX (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.
XX Kim SY;
XX WPI; 2001-217937/22.
XX
```

```
XX New nucleic acid molecule encoding Abscisic acid responsive element
PT binding factor 1 (ABF1) which can be used to generate transgenic plants
PT that are tolerant to multiple environmental stresses
XX
XX Disclosure; Fig 8A; 42pp; English.
XX
XX The sequence represents the amino acid sequence of basic leucine zipper
CC (bZIP) region abscisic acid responsive element binding factor 3 (ABF3),
CC isolated from an Arabidopsis cDNA library. Abscisic acid (ABA) is
CC a major plant hormone involved in response to adverse environmental
CC conditions such as drought, high salt and cold/freezing. This response
CC involves induced expression of various genes. ABA-responsive elements
CC (ABREs) are cis-regulatory elements that mediate the ABA-modulated gene
CC expression, and interact with a novel class of ABRE-binding factors
CC (ABFs). ABFs are basic leucine zipper (bZIP) class transcription factors
CC that bind to both G/ABREs and C/ABREs. ABFs have the potential to
CC activate a large number of ABA/stress responsive genes and can be used to
CC generate transgenic plants that are tolerant to multiple environmental
CC stresses.
XX
XX Sequence 60 AA;
SQ
XX
XX Query Match 4.1%; Score 60; DB 22; Length 60;
Best Local Similarity 29.5%; Pred. No. 9.3e+02;
Matches 13; Conservative 13; Mismatches 18; Indels 0; Gaps 0;
QY 136 KQLIADPKNKETYEKNLKAIVAKLEKDKAKSPDAIAENKK 179
Db 16 RESAARSRARKQAVTMELEAIEIQLKELNEELQKKQVEIMEKK 59
XX
XX RESULT 14
XX AAB53983
XX ID AAB53983 standard; Protein; 58 AA.
XX
XX AAB53983;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:1523.
XX
XX Human; colon cancer; colon cancer antigen; diagnostic; detection;
XX immunofluorescence; cytochemical; cardiotactile; neuroprotective; vulnery;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX nephrotoxic; anti-infective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX reproductive disorder; gastrointestinal disorder; renal disorder;
XX infectious disease; cardiovascular disorder.
XX
XX Homo sapiens.
XX
XX WO200055351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587534/55.
XX
XX N-PSDB; AAC98740.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer
XX
XX Claim 11; Page 2079; 2104pp; English.
XX
```

07-MAY-1999; 99US-0132863.

PR	14-MAY-1999;	99US-0134211
PR	14-MAY-1999;	99US-0134221
PR	14-MAY-1999;	99US-0134221
PR	14-MAY-1999;	99US-0134370
PR	18-MAY-1999;	99US-0134766
PR	19-MAY-1999;	99US-0134944
PR	20-MAY-1999;	99US-0135124
PR	21-MAY-1999;	99US-0135355
PR	25-MAY-1999;	99US-0136021
PR	27-MAY-1999;	99US-0136282
PR	28-MAY-1999;	99US-0136782
PR	01-JUN-1999;	99US-0137222
PR	03-JUN-1999;	99US-0137558
PR	04-JUN-1999;	99US-0137502
PR	07-JUN-1999;	99US-0137754
PR	08-JUN-1999;	99US-0138094
PR	10-JUN-1999;	99US-0138540
PR	10-JUN-1999;	99US-0138847
PR	14-JUN-1999;	99US-0139119
PR	16-JUN-1999;	99US-0139452
PR	16-JUN-1999;	99US-0139453
PR	17-JUN-1999;	99US-0139492
PR	18-JUN-1999;	99US-0139444
PR	18-JUN-1999;	99US-0139445
PR	18-JUN-1999;	99US-0139455
PR	18-JUN-1999;	99US-0139477
PR	18-JUN-1999;	99US-0139487
PR	18-JUN-1999;	99US-0139495
PR	18-JUN-1999;	99US-0139460
PR	18-JUN-1999;	99US-0139461
PR	18-JUN-1999;	99US-0139462
PR	23-JUN-1999;	99US-0140353
PR	23-JUN-1999;	99US-0140354
PR	24-JUN-1999;	99US-0140695
PR	28-JUN-1999;	99US-0140823
PR	30-JUN-1999;	99US-0140991
PR	01-JUL-1999;	99US-0141287
PR	01-JUL-1999;	99US-0141842
PR	02-JUL-1999;	99US-0142154
PR	06-JUL-1999;	99US-0142055
PR	08-JUL-1999;	99US-0142390
PR	08-JUL-1999;	99US-0142803
PR	09-JUL-1999;	99US-0142820
PR	12-JUL-1999;	99US-0142877
PR	13-JUL-1999;	99US-0143524
PR	14-JUL-1999;	99US-0143642
PR	15-JUL-1999;	99US-0144005
PR	16-JUL-1999;	99US-0144085
PR	19-JUL-1999;	99US-0144334
PR	20-JUL-1999;	99US-0144352
PR	20-JUL-1999;	99US-0144632
PR	20-JUL-1999;	99US-0144684
PR	21-JUL-1999;	99US-0144814
PR	21-JUL-1999;	99US-0145086
PR	21-JUL-1999;	99US-0145088
PR	22-JUL-1999;	99US-0145087
PR	22-JUL-1999;	99US-0145089
PR	23-JUL-1999;	99US-0145192
PR	23-JUL-1999;	99US-0145218

PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-01452913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156599.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 57.5; DB 21; Length 59;  
Best Local Similarity 32.3%; Pred. No. 1.5e+03;  
Matches 20; Conservative 10; Mismatches 15; Indels 17; Gaps 3;

QY 126 NGIITSKNIKQLIADPKKKEIYKNIKAYVAKLEIKDEAKSKFPAIAENKLIYTSE 185  
DB 3 NGIARSPSIKK---AKNINSKSYAK-----EATDEWARF-----AKKELLIRSS 45  
QY 186 GC 187  
DB 46 GC 47

Search completed: February 11, 2004, 16:49:40  
Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:48:59 ; Search time 34 Seconds  
(without alignments)  
1767.433 Million cell updates/sec

Title: US-09-869-677a-2  
Perfect score: 1475  
Sequence: 1 STGAKTKAKSDKLKLVATNS.....PDSYVAMKMLDKISEGL 287

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 266164

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PTC\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/PTCUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	8.1	27	US-09-027-956-7	Sequence 7, Appl1
2	62	4.2	56	US-09-816-989A-3	Sequence 3, Appl1
3	58	3.9	58	US-09-925-299-1523	Sequence 1523, Ap
4	58	3.9	58	US-09-925-299-1523	Sequence 1523, Ap
5	53.5	3.6	59	US-09-764-869-772	Sequence 772, App
6	53.5	3.6	59	US-10-427-577-772	Sequence 772, App
7	53.5	3.6	59	US-10-091-504-772	Sequence 772, App
8	53.5	3.6	59	US-10-091-504-772	Sequence 772, App
9	53	3.6	52	US-10-383-675-11	Sequence 11, Appl
10	51.5	3.5	37	US-09-864-761-44177	Sequence 44177, A
11	51.5	3.5	54	US-10-161-660-23	Sequence 23, Appl
12	51.5	3.5	59	US-09-764-891-3185	Sequence 3185, Ap
13	51.5	3.5	59	US-10-205-428-356	Sequence 356, App
14	50	3.4	51	US-09-864-761-45990	Sequence 45990, A
15	50	3.4	52	US-09-820-843A-63	Sequence 63, Appl

16	49	3.3	16	US-09-351-794A-7	Sequence 7, Appl1
17	49	3.3	50	US-10-029-386-34189	Sequence 34189, A
18	48.5	3.3	54	US-10-080-505-38	Sequence 38, Appl
19	48	3.3	37	US-10-357-884-16	Sequence 16, Appl
20	48	3.3	44	US-09-864-761-41610	Sequence 41610, A
21	48	3.3	50	US-09-864-761-48301	Sequence 48301, A
22	48	3.3	52	US-09-864-408A-3108	Sequence 3108, App
23	48	3.3	55	US-09-764-846-265	Sequence 265, App
24	48	3.3	55	US-10-091-483-265	Sequence 265, App
25	47.5	3.2	35	US-10-299-043-5	Sequence 5, Appl1
26	47.5	3.2	44	US-09-864-761-47883	Sequence 47883, A
27	47	3.2	56	US-10-023-066A-79	Sequence 79, Appl
28	47	3.2	58	US-10-178-213-228	Sequence 228, App
29	47	3.2	60	US-10-395-740-14	Sequence 14, Appl
30	46.5	3.2	32	US-09-864-761-37669	Sequence 37669, A
31	46.5	3.2	41	US-09-864-761-38343	Sequence 38343, A
32	46.5	3.2	43	US-10-321-857-36	Sequence 36, Appl
33	46.5	3.2	43	US-10-318-675-36	Sequence 36, Appl
34	46.5	3.2	48	US-09-864-761-41071	Sequence 41071, A
35	46	3.1	45	US-09-813-153-203	Sequence 203, App
36	46	3.1	50	US-09-864-761-45713	Sequence 45713, A
37	46	3.1	51	US-09-757-049A-3	Sequence 3, Appl1
38	46	3.1	54	US-10-300-072-45	Sequence 45, Appl
39	46	3.1	56	US-10-023-066A-77	Sequence 77, Appl
40	46	3.1	58	US-09-933-767-950	Sequence 950, App
41	46	3.1	58	US-10-023-282-950	Sequence 950, App
42	46	3.1	60	US-10-000-256A-225	Sequence 225, App
43	46	3.1	37	US-10-105-232-318	Sequence 318, App
44	45.5	3.1	37	US-10-189-437-305	Sequence 305, App
45	45.5	3.1	47	US-10-105-232-317	Sequence 317, App

# ALIGNMENTS

RESULT 1  
US-09-027-956-7  
Sequence 7, Application US/09027956A  
Patent No. US20010048929A1  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: LINDBERG, Alf  
APPLICANT: KLEIN, Michael  
TITLE OF INVENTION: NOVEL MULTI-OLIGOSACCHARIDE GLYCOCONJUGATE BACTERIAL  
FILE REFERENCE: 1038-791 MTS:jd  
CURRENT APPLICATION NUMBER: US/09/027,956A  
CURRENT FILING DATE: 1998-02-23  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 7  
LENGTH: 27  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-027-956-7

Query Match 8.1% Score 120, DB 9, Length 27;  
Best Local Similarity 88.5% Pred. No. 0.00056;  
Matches 23; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 127 GIYKKNIAKQIAKPKNKETEYKN 152  
1 GIYKKNIAKQIAKPKNKETEYKN 26  
RESULT 2  
US-09-816-989A-3  
Sequence 3, Application US/09816989A  
Patent No. US20020115103A1  
GENERAL INFORMATION:  
APPLICANT: Gad, Alexander  
APPLICANT: Lis, Doris  
TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MAR

```

; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

```

```

Query Match
Best Local Similarity 38.9%; Score 62; DB 10; Length 56;
Matches 21; Conservative 7; Mismatches 22; Indels 4; Gaps 2;

```

```

Cy 149 YEKNIKAYAKLEKDKKSKSPDAIENKKLIVTSBGCRK-YFSKAYGVPSAY 201
Db 4 YAKKKAAYAKKAERAKKAERKAYKAKAAK--KAAKRYKAAKAAKAAKAAAY 54

```

```

RESULT 3
US-09-925-299-1523

```

```

; Sequence 1523, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1523
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1523

```

```

Query Match
Best Local Similarity 34.6%; Score 58; DB 9; Length 58;
Matches 18; Conservative 8; Mismatches 24; Indels 2; Gaps 1;

```

```

Cy 119 HAMINLENGIITSNIAKQIADPDKKETEYKNIKAVATLEKLDKDEAKSK 170
Db 2 HALHLTKNKL--CSNCHKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 51

```

```

RESULT 4
US-09-925-299-1523
; Sequence 1523, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:

```

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1523
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1523

```

```

Query Match
Best Local Similarity 34.6%; Score 58; DB 11; Length 58;
Matches 18; Conservative 8; Mismatches 24; Indels 2; Gaps 1;

```

```

Cy 119 HAMINLENGIITSNIAKQIADPDKKETEYKNIKAVATLEKLDKDEAKSK 170
Db 2 HALHLTKNKL--CSNCHKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 51

```

```

RESULT 5
US-09-764-869-772

```

```

; Sequence 772, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 772
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (40)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-772

```

```

Query Match
Best Local Similarity 29.7%; Score 53.5; DB 9; Length 59;
Matches 11; Conservative 9; Mismatches 16; Indels 1; Gaps 1;

```

```

Cy 237 YARRMEIVSKDSIGIPYSEIFTDISIAKGGKGDSTY 273
Db 14 VASKPLVTSNPGGLP-YMKVFDTLGDGDTGTTTF 49

```

```

RESULT 6
US-10-227-577-772
; Sequence 772, Application US/10227577
; Publication No. US20040005575A1
; GENERAL INFORMATION:

```

```
APPLICANT: Rosen et al.
FILE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C2
CURRENT APPLICATION NUMBER: US/10/227,577
PRIOR APPLICATION NUMBER: 10/091,504
PRIOR FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 10/091,504
PRIOR FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: 09/764,869
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 772
LENGTH: 59
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-227-577-772

Query Match
Best Local Similarity 3.6%; Score 53.5; DB 12; Length 59;
Matches 11; Conservative 9; Mismatches 16; Indels 1; Gaps 1;

QY 237 VDRPMEVSKDSGIPYSEIFTSIAKKGKPGDSY 273
Db 14 VKSKPLVSISSNPGSLP-YMKVFDTLGLXDMWTGTTP 49

RESULT 7
US-10-091-504-772
; Sequence 772, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 772
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-091-504-772

Query Match
Best Local Similarity 3.6%; Score 53.5; DB 15; Length 59;
Matches 11; Conservative 9; Mismatches 16; Indels 1; Gaps 1;
```

```
QY 237 VDRPMEVSKDSGIPYSEIFTSIAKKGKPGDSY 273
Db 14 VKSKPLVSISSNPGSLP-YMKVFDTLGLXDMWTGTTP 49

RESULT 8
US-09-730-763-11
; Sequence 11, Application US/09730763
; Publication No. US20020192781A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, Mary
; APPLICANT: GIOUET, Brigitte
; TITLE OF INVENTION: DESATURASE ANTIGEN OF MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: c/o Finnegan Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/730,763
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/917,299
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MEYERS, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0156-00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4400
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-730-763-11

Query Match
Best Local Similarity 3.6%; Score 53; DB 10; Length 52;
Matches 14; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 168 KSKFDALIAENKKLIVTSGCFKTFYSKAYG 196
Db 4 KSHYGAIAVGEVAAVVGEGMARFSPKPG 32

RESULT 9
US-10-383-675-11
; Sequence 11, Application US/10383675
; Publication No. US20030157589A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, Mary
; APPLICANT: Gicquel, Brigitte
; TITLE OF INVENTION: METHOD OF SCREENING ANTI-MYCOBACTERIAL MOLECULES
; FILE REFERENCE: 03495.0182-00000
; CURRENT APPLICATION NUMBER: US/10/383,675
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: US/09/429,370
```

TYPE: PRT

```
query match {
  3.5* score 51.5; db 12; length 54;
```



Best Local Similarity 31.4%; Pred. No. 2e+03;  
Matches 16; Conservative 11; Mismatches 13; Indels 11; Gaps 4;

QY 81 TKLVKMNQKTKNDYFAV-----SDGDIVYLLGASAKGKEDHAMLNTLE 125  
Db 6 TKVTFHVEKSNPNYLLALVRYVNDG-DVVAVD-IREKGD---KWTLEK 51

## RESULT 12

US-09-764-891-3185  
; Sequence 3185, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; Prior Filing Date: 2001-01-17  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3185  
; LENGTH: 59  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (2)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (12)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (58)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-3185

Query Match 3.5%; Score 51.5; DB 11; Length 59;

Best Local Similarity 30.2%; Pred. No. 2.3e+03;  
Matches 13; Conservative 7; Mismatches 12; Indels 11; Gaps 1;

QY 102 IDVYLLGASEKGEKEDP-----HAWMLNENGIYISKN 133  
Db 5 LTVLYQDXAHCKSEKPSLMGYIMRMIMHRSASVYDGMSSKD 47

## RESULT 13

US-10-205-428-356  
; Sequence 356, Application US/10205428  
; Publication No. US20030108907A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P0117C1  
; CURRENT APPLICATION NUMBER: US/10/205,428  
; Prior Filing Date: 2002-07-26  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447

; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1019  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 356  
; LENGTH: 59  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (2)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (12)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (58)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-205-428-356

Query Match 3.5%; Score 51.5; DB 15; Length 59;

Best Local Similarity 30.2%; Pred. No. 2.3e+03;  
Matches 13; Conservative 7; Mismatches 12; Indels 11; Gaps 1;

QY 102 IDVYLLGASEKGEKEDP-----HAWMLNENGIYISKN 133  
Db 5 LTVLYQDXAHCKSEKPSLMGYIMRMIMHRSASVYDGMSSKD 47

## RESULT 14

US-09-864-761-45990  
; Sequence 45990, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30

Job time : 41 secs

```
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 45990
/ LENGTH: 51
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AL121694.3
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
/ OTHER INFORMATION: SWISSPROT HIT: 055131, EVALU 2.70e+00
/ US-09-864-761-45990
```

```
Query Match 3.4%; Score 50; DB 9; Length 51;
Best Local Similarity 23.4%; Pred. No. 2.5e+03;
Matches 15; Conservative 9; Mismatches 18; Indels 22; Gaps 1;
```

```
OY 84 VNMACTKQKDYFAVSDIDVYTLGASBKEDPHAWLNLENGIYSKNIKQIADKP 143
Db 10 LRDQSKQKDH-----ETLKNLENGFYINDQKADLLLENK 47
OY 144 KNKE 147
Db 48 KLKE 51
```

```
RESULT 15
US-09-820-843A-63
/ Sequence 63, Application US/09820843A
/ Publication No. US20030039963A1
/ GENERAL INFORMATION:
/ APPLICANT: Council of Scientific and Industrial Research
/ TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
/ TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
/ FILE REFERENCE: Q63915
/ CURRENT APPLICATION NUMBER: US/09/820,843A
/ CURRENT FILING DATE: 2001-03-30
/ NUMBER OF SEQ ID NOS: 118
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 63
/ LENGTH: 52
/ TYPE: PRT
/ ORGANISM: H. influenzae
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: predicted coding region H1562
/ NAME/KEY: misc feature
/ OTHER INFORMATION: g1|1574414
/ US-09-820-843A-63
```

```
Query Match 3.4%; Score 50; DB 11; Length 52;
Best Local Similarity 28.3%; Pred. No. 2.6e+03;
Matches 15; Conservative 10; Mismatches 16; Indels 12; Gaps 2;
```

```
OY 138 LIAKPKKETEYENLKAAYAKLEKDKAKSPDAIENKLIIVTSGGCFKY 190
Db 1 MLSKDPK-----VLIKGELEKDKSKAKKYPGDACD---LRSGGCDKY 41
```

Search completed: February 11, 2004, 16:52:27

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:50:34 ; Search time 23 Seconds

(without alignments)  
962.039 Million cell updates/sec

Title: US-09-869-677A-2

Perfect score: 1475  
Sequence: 1 SSTGAKTKAKSDKLKVATNS.....PGDSYAMMKMNDKISEGL 287

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 378584 seqs, 7869973 residues

Total number of hits satisfying chosen parameters: 181738

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54.5	3.7	60	US-09-712-638A-10	Sequence 10, Appl
2	53	3.6	52	US-10-383-675A-11	Sequence 11, Appl
3	51	3.5	52	US-10-390-472-8	Sequence 8, Appl
4	50	3.4	44	PCT-US03-23249-7	Sequence 7, Appl
5	50	3.4	52	US-09-820-843B-63	Sequence 63, Appl
6	49	3.3	53	PCT-US03-09632-142	Sequence 142, Appl
7	49	3.3	59	US-60-498-170-10	Sequence 10, Appl
8	49	3.3	60	US-09-714-602-357	Sequence 357, Appl
9	48.5	3.3	51	US-09-711-161A-737	Sequence 737, Appl
10	48.5	3.3	54	US-10-687-046-38	Sequence 38, Appl
11	48	3.3	48	PCT-US02-41612A-461	Sequence 461, Appl
12	47	3.2	56	US-10-425-114A-40154	Sequence 40154, A
13	47	3.2	60	PCT-US03-09572A-14	Sequence 14, Appl
14	46	3.1	45	US-10-631-141-1853	Sequence 1853, Ap
15	46	3.1	45	US-10-631-441A-1853	Sequence 1853, Ap
16	46	3.1	46	US-09-897-516A-7904	Sequence 7904, Ap
17	46	3.1	49	PCT-US02-3760-35	Sequence 35, Appl
18	46	3.1	54	PCT-US03-37507-45	Sequence 45, Appl
19	46	3.1	54	US-10-456-947-30	Sequence 30, Appl
20	46	3.1	54	US-10-717-984-45	Sequence 45, Appl
21	46	3.1	59	US-09-897-516A-8227	Sequence 8227, Appl
22	46	3.1	59	PCT-US03-09632-72	Sequence 72, Appl
23	45.5	3.1	57	US-10-425-114A-59888	Sequence 59888, A
24	45.5	3.1	60	US-10-631-141-3040	Sequence 3040, Ap
25	45.5	3.1	60	US-10-631-441A-3040	Sequence 3040, Ap
26	45.5	3.1	60	US-10-631-441A-3040	Sequence 3040, Ap

27	45	3.1	38	US-10-415-182A-8032	Sequence 8032, Ap
28	45	3.1	57	US-10-631-141-2890	Sequence 2890, Ap
29	45	3.1	57	US-10-631-441A-2890	Sequence 2890, Ap
30	45	3.1	58	US-09-897-516A-5827	Sequence 5827, Ap
31	44	3.0	35	US-09-879-666D-5	Sequence 5, Appl
32	44	3.0	35	US-10-632-886A-5	Sequence 5, Appl
33	44	3.0	50	PCT-US03-38808-99	Sequence 99, Appl
34	44	3.0	50	US-10-639-067-200	Sequence 200, Appl
35	44	3.0	58	US-10-664-025-4354	Sequence 4354, Ap
36	44	3.0	58	US-10-644-025-4354	Sequence 4354, Ap
37	43.5	2.9	60	US-10-472-928-4476	Sequence 4476, Ap
38	43	2.9	19	US-09-392-941B-132	Sequence 132, Appl
39	43	2.9	19	US-09-392-941B-132	Sequence 132, Appl
40	43	2.9	19	US-09-392-941B-132	Sequence 132, Appl
41	43	2.9	5	US-10-415-182A-8460	Sequence 164, Appl
42	43	2.9	41	US-10-664-025-5996	Sequence 8460, Ap
43	43	2.9	53	US-10-664-025-5996	Sequence 5996, Ap
44	43	2.9	56	US-10-473-551-1122	Sequence 1122, Ap
45	43	2.9	60	PCT-US02-41612A-420	Sequence 420, Appl

## ALIGNMENTS

```
RESULT 1
US-09-712-638A-10
Sequence 10, Application US/09712638A
GENERAL INFORMATION:
APPLICANT: IBM Corporation
TITLE OF INVENTION: Unsuperised Building and Exploitation of Composite Descriptors
FILE REFERENCE: YOR92000435US1
CURRENT APPLICATION NUMBER: US/09/712.638A
CURRENT FILING DATE: 2000-11-14
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatentIn version 3.2
SEQ ID NO 10
LENGTH: 60
TYPE: PRT
ORGANISM: Fugu rubripes
US-09-712-638A-10

Query Match          3.7%; Score 54.5; DB 5; Length 60;
Best Local Similarity 41.7%; Pred. No. 5.1e+03;
Matches 15; Conservative 2; Mismatches 12; Indels 7; Gaps 1;

QY      150 EKNIKAYVAFLEKIDK-----EAKSKFDALDENK 178
Db       3 EAOLKEAKKREKLEKFKQKKEWEAKKKQGPVAAKK 38

RESULT 2
US-10-383-675A-11
Sequence 11, Application US/10383675A
GENERAL INFORMATION:
APPLICANT: JACKSON, MARY
APPLICANT: GICQUEL, BRIGITTE
TITLE OF INVENTION: METHOD OF SCREENING ANTI-MYCOCBACTERIAL MOLECULES
FILE REFERENCE: 03495.0182-01
CURRENT APPLICATION NUMBER: US/10/383.675A
CURRENT FILING DATE: 2003-03-10
PRIOR APPLICATION NUMBER: 09/429,370
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/113,375
PRIOR FILING DATE: 1998-11-04
PRIOR APPLICATION NUMBER: 60/111,813
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/181,934
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 52
TYPE: PRT
```

ORGANISM: Pseudomonas mendocina  
US-10-383-675A-11

Query Match 3.6%; Score 53; DB 6; Length 52;  
Best Local Similarity 48.3%; Pred. No. 5.5e+03;  
Matches 14; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

Qy 168 KSKPDAIENKGLIVSEGCFFKFSYAYG 196  
Db 4 KSHYGAIWGEVAAYVTEGRMARFSKAPG 32

RESULT 3  
US-10-390-472-8

Sequence 8, Application US/10390472  
GENERAL INFORMATION:

APPLICANT: Holmes, Todd

Zhang, Shuangang  
Rich, Alexander  
DiPersio, C. Michael  
Lockshin, Curtis

TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY  
SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES  
THEREFOR

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02173-4799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/390,472

FILING DATE: 17-Mar-2003

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/293,284

FILING DATE: 22-AUG-1994

APPLICATION NUMBER: 07/973,326

FILING DATE: 28-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: MIT-6008A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 52 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-10-390-472-8

Query Match 3.5%; Score 51; DB 6; Length 52;  
Best Local Similarity 26.1%; Pred. No. 7.1e+03;  
Matches 12; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

RESULT 4  
PCT-US03-23249-7

Sequence 7, Application PC/TUS0323249  
GENERAL INFORMATION:

APPLICANT: INCYTE CORPORATION

ELIOTT, Vicki S.; SWANNARAR, Anita;

GRIFFIN, Jennifer A.; LEE, Ernestine A.;

SPRAGUE, William W.; HAFALIA, April J.A.;

LEE, Soo Yeun; KHARE, Amy E.;

ISON, Craig H.; KHARE, Reena;

CHAWLA, Marinder K.; MARQUIS, Joseph P.;

JIANG Xin; JACKSON, Alan A.;

BECHA, Shanya D.; EMERLING, Brooke M.;

JIN, Pei; WILSON, Amy D.;

RICHARDSON, Thomas W.; YANG, Junning;

BAUGHN, Mariah R.; GANDHI, Ameena R.;

NGUYEN, Daniel B.; RAMKUMAR, Jayalaksh;

KALLICK, Deborah A.; KEARNEY, Liam;

LU, Dyung Anna M.; GRETZEN, Kimberly J.;

TRIBOLEY, Catherine M.; LAU, Preeti G.;

BLAKE, Julie J.; LU, Yan;

APPLICANT: ARYZO, Chandra S.

TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

FILE REFERENCE: PF-1478 PCT

CURRENT APPLICATION NUMBER: PCT/US03/23249

CURRENT FILING DATE: 2003-07-23

PRIOR APPLICATION NUMBER: US 60/398,143

PRIOR FILING DATE: 2002-07-23

PRIOR APPLICATION NUMBER: US 60/402,458

PRIOR FILING DATE: 2002-08-09

PRIOR APPLICATION NUMBER: US 60/403,289

PRIOR FILING DATE: 2002-08-12

PRIOR APPLICATION NUMBER: US 60/406,472

PRIOR FILING DATE: 2002-08-27

PRIOR APPLICATION NUMBER: US 60/409,354

PRIOR FILING DATE: 2002-09-06

NUMBER OF SEQ ID NOS: 124

SOFTWARE: PERL Program

SEQ ID NO: 7

LENGTH: 44

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No: 7512761CD1

PCT-US03-23249-7

Query Match 3.4%; Score 50; DB 1; Length 44;  
Best Local Similarity 33.3%; Pred. No. 7.1e+03;  
Matches 11; Conservative 5; Mismatches 15; Indels 2; Gaps 1;

Qy 106 YLGASEKEDPHAMNLENGITTSKNTAKOL 138  
Db 7 FLOOSGSSGPGQPAVWHRLE--LYTKNFMKTL 37

RESULT 5  
US-09-820-843B-63

Sequence 63, Application US/09820843B  
GENERAL INFORMATION:

APPLICANT: Council of Scientific and Industrial Research

TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE

FILE REFERENCE: 063915

CURRENT APPLICATION NUMBER: US/09/820,843B

CURRENT FILING DATE: 2001-03-30

NUMBER OF SEQ ID NOS: 118

SOFTWARE: Patentin version 3.2

SEQ ID NO: 63

LENGTH: 52

TYPE: PRT

ORGANISM: H. influenzae

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: predicted coding region H11562

FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: gi|1574414  
US-09-820-843B-63

Query Match 3.4%; Score 50; DB 5; Length 52;  
Best Local Similarity 28.3%; Pred. No. 8e+03;  
Matches 15; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

QY 138 LINKDPKNETYEKLNKAYAKLEKLDKSKSKDALKENKCLIVTSEGCPEKY 190  
Db 1 MLSKDPK-----VLKLGELKSKSKAKKXFGDAD----LRSGEGCDKY 41

RESULT 6  
PCT-US03-09632-142  
Sequence 142, Application PC/TUS0309632  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL TWO SPLICED VARIANTS OF A HUMAN  
TITLE OF INVENTION: CELL SURFACE PROTEIN WITH IMMUNOLOGICAL FOLDS, BGSSG AND BGSS1  
FILE REFERENCE: D0228 PCT  
CURRENT APPLICATION NUMBER: PCT/US03/09632  
CURRENT FILING DATE: 2003-03-28  
PRIOR APPLICATION NUMBER: U.S. 60/368,671  
PRIOR FILING DATE: 2002-03-29  
PRIOR APPLICATION NUMBER: U.S. 60/371,420  
PRIOR FILING DATE: 2002-04-10  
NUMBER OF SEQ ID NOS: 156  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 142  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US03-09632-142

Query Match 3.3%; Score 49; DB 1; Length 53;  
Best Local Similarity 33.3%; Pred. No. 9.3e+03;  
Matches 12; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

QY 27 KALAGKIDILHSIVPIGQDPHEYEPLPEDAKETSNA 62  
Db 16 QAVGVGLIELHCEAPRGSPILWVFYHEDVTLGSSS 51

RESULT 7  
US-60-498-170-10  
Sequence 10, Application US/60498170  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING A NOVEL HUMAN NEURONAL CELL ADHESION  
TITLE OF INVENTION: PROTEIN, BGS-28, AND VARIANTS THEREOF  
FILE REFERENCE: D0245 PSP  
CURRENT APPLICATION NUMBER: US/60/498,170  
CURRENT FILING DATE: 2003-08-27  
NUMBER OF SEQ ID NOS: 77  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 10  
LENGTH: 59  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-60-498-170-10

Query Match 3.3%; Score 49; DB 7; Length 59;  
Best Local Similarity 35.1%; Pred. No. 1e+04;  
Matches 13; Conservative 3; Mismatches 13; Indels 8; Gaps 1;

QY 40 VPIGDPHEYEPLPEDAKETSNAIVFYNGINLEDDG 76  
Db 25 VPIIDTDERYTLPLPK-----GVLDITSLRADGG 53

RESULT 8

US-09-714-602-357

Sequence 357, Application US/09714602  
GENERAL INFORMATION:

APPLICANT: David William Holden

TITLE OF INVENTION: Identification of Genes

NUMBER OF SEQUENCES: 501

CORRESPONDENCE ADDRESS:

ADDRESSEE: Patrea L. Pabst

STREET: 2800 One Atlantic Center

1201 West Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30309-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/714,602

FILING DATE: 16-Nov-2000

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/871,355A

FILING DATE: 09-JUN-1997

APPLICATION NUMBER: PCT/GB95/02875

FILING DATE: 11-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: RPMS 101 CON

TELEPHONE: (404) 873-8794

TELEFAX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 357:

SEQUENCE CHARACTERISTICS:

LENGTH: 60 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 357:

US-09-714-602-357

Query Match 3.3%; Score 49; DB 5; Length 60;  
Best Local Similarity 24.6%; Pred. No. 1e+04;  
Matches 16; Conservative 9; Mismatches 20; Indels 20; Gaps 2;

QY 49 YEPLPEDAKETSNAIVFYNGINLEDDGQAMFTLVKNAOKTKQDYFAVSDGIDVLYLE 108  
Db 11 FEQLLDEASTSNA-----SALLPQVWLTRQWDMQMLTVGD--YLA 50

QY 109 GASEK 113  
Db 51 RISRR 55

RESULT 9

US-09-711-161A-737  
Sequence 737, Application US/09711161A  
GENERAL INFORMATION:

APPLICANT: LEHRER, SAMUEL B.

APPLICANT: REESE, GERRARD

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODIFYING AN IMMUNE

TITLE OF INVENTION: RESPONSE AGAINST TROPOMYOSIN

FILE REFERENCE: 55394(45406)

CURRENT APPLICATION NUMBER: US/09/711,161A

CURRENT FILING DATE: 2000-11-12

PRIOR APPLICATION NUMBER: 60/165,226

PRIOR FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 1150

```
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 737
LENGTH: 51
TYPE: PRT
ORGANISM: Dermatophagoides pteronyssinus
US-09-711-161A-737
```

```
Query Match 3.3%; Score 48.5; DB 5; Length 51;
Best Local Similarity 32.7%; Pred. No. 9.6e+03;
Matches 17; Conservative 7; Mismatches 25; Indels 3; Gaps 1;
```

```
Qy 116 EDPHAMLNLENGIYISKNIKQIAKDPKNEYKLNKAYAKLEKDKEA 167
Db 3 EEKEKALQTAGDVAAALNRRIQLIED--LERSEERKIKATKEASQSA 51
```

```
RESULT 10
US-10-687-046-38
Sequence 38, Application US/10687046
GENERAL INFORMATION:
APPLICANT: St. Gene, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RT/DCF/DHR
CURRENT FILING DATE: 2003-10-15
PRIOR APPLICATION NUMBER: US/10/080,505
PRIOR FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 38
LENGTH: 54
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-687-046-38
```

```
Query Match 3.3%; Score 48.5; DB 6; Length 54;
Best Local Similarity 25.5%; Pred. No. 1e+04;
Matches 13; Conservative 13; Mismatches 20; Indels 5; Gaps 2;
```

```
Qy 161 EKLDEKAKSKPDIAENKLIIVTSEGCPSKAYGVPSAYIWEINTEBEG 211
Db 7 ETTTSAHRRNTLVNKL--SGCGTFQPTSSIFGKSD---KLSDABG 52
```

```
RESULT 11
PCT-US02-41612A-461
Sequence 461, Application PC/TUS0241612A
GENERAL INFORMATION:
APPLICANT: diadexus, Inc.
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Endometrial Specific Genes
FILE REFERENCE: DEX-0378
CURRENT APPLICATION NUMBER: PCT/US02/41612A
CURRENT FILING DATE: 2003-06-17
PRIOR APPLICATION NUMBER: 60/342,756
PRIOR FILING DATE: 2001-12-21
NUMBER OF SEQ ID NOS: 600
SOFTWARE: PatentIn version 3.1
SEQ ID NO 461
LENGTH: 48
TYPE: PRT
ORGANISM: Homo sapien
PCT-US02-41612A-461
```

```
Query Match 3.3%; Score 48; DB 1; Length 48;
Best Local Similarity 28.3%; Pred. No. 9.7e+03;
Matches 15; Conservative 13; Mismatches 9; Indels 16; Gaps 4;
```

```
Qy 176 ENK-KLIVTSEC---CRKPSKAYGVPSAYIWEINTEBEGTPQISLIFLK 224
Db 2 ENRGKIVHTSGNKLKKRYN-----IWLHSE--PGLNLTIERMK 42
```

```
RESULT 12
US-10-425-114A-40154
Sequence 40154, Application US/10425114A
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21 (5313) B
CURRENT APPLICATION NUMBER: US/10/425,114A
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 40154
LENGTH: 56
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 701046730_FLI.pep
US-10-425-114A-40154
```

```
Query Match 3.2%; Score 47; DB 6; Length 56;
Best Local Similarity 45.5%; Pred. No. 1.2e+04;
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy 139 IAKDPKNEYKLNKAYAKL 160
Db 22 VATAEKKQGEYKTLIAYNKKL 43
```

```
RESULT 13
PCT-US03-09572A-14
Sequence 14, Application PC/TUS0309572A
GENERAL INFORMATION:
APPLICANT: WYETH
APPLICANT: Mark, Robert
APPLICANT: Wood, Andrew
APPLICANT: Gulukota, Kamalakart
TITLE OF INVENTION: NOVEL PANCORLIN-PABLO PROTEIN INTERACTIONS AND METHODS OF USE
FILE REFERENCE: AM100375PCT
CURRENT APPLICATION NUMBER: PCT/US03/09572A
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: US 60/369,244
PRIOR FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: US 60/386,645
PRIOR FILING DATE: 2002-06-06
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.2
SEQ ID NO 14
LENGTH: 60
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US03-09572A-14
```

```
Query Match 3.2%; Score 47; DB 1; Length 60;
Best Local Similarity 29.4%; Pred. No. 1.3e+04;
Matches 15; Conservative 11; Mismatches 17; Indels 8; Gaps 2;
```

```
Qy 136 KQIAKDPKNEYKLNKAYAKLEKDKAKSKPDIAENK 179
Db 2 RQLEKRYQMSQSIETVLDRTORDIQ-YEKKENQMKGLSEKFKQVEBSHK 51
```

```
RESULT 14
US-10-631-141-1853
; Sequence 1853, Application US/10631141
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; FILE REFERENCE: GEN-T119C1
; CURRENT APPLICATION NUMBER: US/10/631,141
; CURRENT FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: US/09/547,599C
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 08/905,223
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,135
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,051
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,144
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,279
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/904,468
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,134
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,133
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 3475
; SOFTWARE: Patent.pm
; SEQ ID NO 1853
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: Brain
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -13...1
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.10
; OTHER INFORMATION: seq MLILSONIAQLEA/QV
US-10-631-141-1853

Query Match      3.1%; Score 46; DB 6; Length 45;
Best Local Similarity 34.5%; Pred. No. 1.2e+04;
Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

OY      151  KNLKAVYAKLEKDKPKAKSKPDALAEKXK 179
Db      6  QNIAQLERQVEKTKKISAINQLENSK 34

RESULT 15
US-10-631-441A-1853
; Sequence 1853, Application US/10631141A
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; FILE REFERENCE: GEN-T119C1
; CURRENT APPLICATION NUMBER: US/10/631,441A
; CURRENT FILING DATE: 2003-07-30
; PRIOR APPLICATION NUMBER: US/09/547,599C
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 08/905,223
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,135
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,051
; PRIOR FILING DATE: 1997-08-01
```

```
; PRIOR APPLICATION NUMBER: US 08/905,144
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,279
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/904,468
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,134
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,133
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 3475
; SOFTWARE: Patent.pm
; SEQ ID NO 1853
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: Brain
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -13...1
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.10
; OTHER INFORMATION: seq MLILSONIAQLEA/QV
US-10-631-441A-1853

Query Match      3.1%; Score 46; DB 6; Length 45;
Best Local Similarity 34.5%; Pred. No. 1.2e+04;
Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

OY      151  KNLKAVYAKLEKDKPKAKSKPDALAEKXK 179
Db      6  QNIAQLERQVEKTKKISAINQLENSK 34
```

Search completed: February 11, 2004, 16:56:03  
Job time : 29 secs

**THIS PAGE BLANK (USPTO)**